Université de Montréal

Genes involved in the metabolism of fatty acids and risk for Crohn's disease in children: A candidate gene study

Par

Irina C. COSTEA

Département de Médecine Sociale et Préventive
Faculté de Médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en Santé Publique
Option Épidémiologie

Février, 2009

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Genes involved in the metabolism of fatty acids and risk for Crohn's disease in children: A candidate gene study

présentée par:
Irina C. Costea

A été évaluée par un jury compose des personnes suivantes:
Dr Lise Goulet, présidente-rapporteuse
Dr. Devendra Amre, directeur de recherche
Dr. Isabel Fortier, codirectrice
Dr. Marie-Pierre Dubé, membre du jury
Dr. James Engert, examinateur externe
Dr. John David Rioux, représentant du doyen de la FES

Résume

Contexte - La prévalence de la maladie de Crohn (MC), une maladie inflammatoire chronique du tube digestif, chez les enfants canadiens se situe parmi les plus élevées au monde. Les interactions entre les réponses immunes innées et acquises aux microbes de l'hôte pourraient être à la base de la transition de l'inflammation physiologique à une inflammation pathologique. Le leucotriène B4 (LTB4) est un modulateur clé de l'inflammation et a été associé à la MC. Nous avons postulé que les principaux gènes impliqués dans la voie métabolique du LTB4 pourrait conférer une susceptibilité accrue à l'apparition précoce de la MC. Dans cette étude, nous avons exploré les associations potentielles entre les variantes de l'ADN des gènes ALOX5 et CYP4F2 et la survenue précoce de la MC. Nous avons également examiné si les gènes sélectionnés montraient des effets parent-d'origine, influençaient les phénotypes cliniques de la MC et s'il existait des interactions gène-gène qui modifieraient la susceptibilité à développer la MC chez l'enfant.

Méthodes – Dans le cadre d'une étude de cas-parents et de cas-témoins, des cas confirmés, leurs parents et des contrôles ont été recrutés à partir de trois cliniques de gastro-entérologie à travers le Canada. Les associations entre les polymorphismes de remplacement d'un nucléotide simple (SNP) dans les gènes CYP4F2 et ALOX5 ont été examinées. Les associations allélique et génotypiques ont été examinées à partir d'une analyse du génotype conditionnel à la parenté (CPG) pour le résultats cas-parents et à l'aide de table de contingence et de régression logistique pour les données de cas-contrôles. Les interactions gène-gène ont été explorées à l'aide de méthodes de réduction multi-factorielles de dimensionnalité (MDR).

Résultats – L'étude de cas-parents a été menée sur 160 trios. L'analyse CPG pour 14 tag-SNP (10 dans la CYP4F2 et 4 dans le gène ALOX5) a révélé la présence d'associations alléliques ou génotypique significatives entre 3 tag-SNP dans le gène CYP4F2 (rs1272, p = 0,04, rs3093158, p = 0.00003, et rs3093145, p = 0,02). Aucune association avec les SNPs de ALOX5 n'a pu être démontrée. L'analyse de l'haplotype de CYP4F2 a montré d'importantes associations avec la MC (test omnibus p = 0,035). Deux haplotypes (GAGTTCGTAA, p = 0,05; GGCCTCGTCG, p = 0,001) montraient des signes d'association avec la MC. Aucun effet parent-d'origine n'a été observé. Les tentatives de réplication pour trois SNPs du gene CYP4F2 dans l'étude cas-témoins comportant 225 cas de MC et 330 contrôles suggèrent l'association dans un de ceux-ci (rs3093158, valeur non-corrigée de p du test unilatéral = 0,03 ; valeur corrigée de p = 0.09). La combinaison des ces deux études a révélé des interactions significatives entre les gènes CYP4F2, ALOX et NOD2. Nous n'avons pu mettre en évidence aucune interaction gène-sexe, de même qu'aucun gène associé aux phénotypes cliniques de la MC n'a pu être identifié.

Conclusions - Notre étude suggère que la CYP4F2, un membre clé de la voie métabolique LTB4 est un gène candidat potentiel pour MC. Nous avons également pu mettre en évidence que les interactions entre les gènes de l'immunité adaptative (CYP4F2 et ALOX5) et les gènes de l'immunité innée (NOD2) modifient les risques de MC chez les enfants. D'autres études sur des cohortes plus importantes sont nécessaires pour confirmer ces conclusions.

Mots-clés: Maladie de Crohn (MC), survenue précoce, gène candidat, CYP4F2, ALOX5, voie de la 5-lipoxygénase (5-LO), leucotriène B4 (LTB4), polymorphismes de remplacement d'un nucléotide simple (SNP).

Abstract

Background - The rates of Crohn's disease (CD) a chronic inflammatory disease of the gastrointestinal tract, among Canadian children are the world's highest. Interactions between the host microbial–innate-immune-responses are thought to underplay transition from physiological to pathological inflammation. Leukotriene B4 (LTB_4) is a key modulator of inflammation and has been shown to be associated with CD. We postulated that key genes involved in the LTB_4 metabolic pathway could confer susceptibility for early-onset CD. In this study we implemented a candidate gene approach to test for associations between DNA variants in the ALOX5 and CYP4F2 genes and early-onset of CD. We also explored whether the selected genes demonstrated parent-of-origin effects, influenced CD clinical phenotypes and whether there were gender-gene and gene-gene interactions that determined CD susceptibility. Methods – The study consisted of an exploratory phase (case-parent design) followed by a replication phase (case-control design). Confirmed cases, parents and controls were recruited from three tertiary gastroenterology clinics across Canada. Associations between tag-single nucleotide polymorphisms in the CYP4F2 and ALOX5 genes were examined. Allelic and/or genotype associations were examined using conditional on parental genotype (CPG) analysis for the case-parent data and contingency table and logistic regression for the case-control data. Gene-gene interactions were explored using multi-factor dimensionality reduction (MDR) methods. Results – The first phase of the study was based on 160 trios (case-parent design). CPG analysis for 14 tag-SNPs (i.e. 10 in the CYP4F2 and 4 in the ALOX5 gene, respectively) revealed significant allelic or genotypic associations between 3 tag-SNPs in the CYP4F2 gene (rs1272, p=0.04, rs3093158, p=0.00003, and rs3093145, p=0.02). No associations with ALOX5 tag-SNPs were evident. CYP4F2-haplotype analysis showed significant associations with CD (omnibus test p-value=0.035). Two specific haplotypes (GAGTTCGTAA, p=0.05; GGCCTCGTCG, p=0.001) showed evidence for association with CD. No parent-of-origin effects were observed. The second phase of the study retested the three CYP4F2 SNPs that showed association in the first stage and was based on 223 CD cases and 330 controls. Some indications of association with one SNP i.e. rs3093158 were present (genotypic uncorrected 1-sided p-value=0.03); however this genotype association did not withstand correction. Combining cases from the two phases of the study revealed significant interactions between the CYP4F2, ALOX and NOD2 genes. No gene-gender interactions were obvious nor were the study genes associated with specific clinical phenotypes of CD. Conclusions - Our study suggests that the CYP4F2, a key member of the LTB_4 metabolic pathway is a potential candidate gene for CD. Furthermore there was evidence that interactions between adaptive immunity genes (CYP4F2 and ALOX5) and innate immunity genes (NOD2) genes modify risk for CD in children. Further studies on larger cohorts are required to confirm these findings.

Keywords: Crohn's Disease (CD), early-onset, candidate gene, CYP4F2, ALOX5, 5-lipoxygenase (5-LO) pathway, leukotriene B4 (LTB4), tag-SNP, case-parent design, case-control design.

Table of content

List of Tables

List of figures

List of Abbreviations

5-LO	5-Lipoxygenase
AA	Arachidonic Acid
ADH	Alcohol dehydrogenase
AIC	Akaike Information Criteria
ALA	Alpha-linolenic Acid
ALDH	aldehyde dehydrogenase
ALOX5	Arachidonate 5-lipoxygenase
ATC	Acute Trauma Center
BCCH	British Columbia's Children Hospital
CD	Crohn's Disease
CDCV	Common disease, common variant
CEPH	Centre d'Etude du Polymorphisme Humain
CHEO	The Children's Hospital of Eastern Ontario
CI	Confidence interval
COR	Case-only odds ratio
CPG	Conditional on parental genotypes
CYP4F2	Cytochrome P450, family 4, subfamily F, polypeptide 2
df	Degree of freedom
DNA	Deoxyribonucleic Acid
DZ	Dizygotic
EM	Expectation-maximization
EPA	Eicosapentaenoic acid
FLAP	5-Lipoxygenase-activating Protein
GIT	Gastro-intestinal Tract
GWA	Genome-wide association
GWAS	Genome-wide association study
GWS	Genome-wide Scan
HSJ	Hopital Ste-Justine
HWE	Hardy-Weinberg Equilibrium
IBD	Inflammatory Bowel Diseases
IL23R	Interleukin 23 Receptor
Km	Specificity constant
LA	Linolenic Acid
LCN	Long-chain
LD	Linkage Disequilibrium
LOD	Logarithm of odds
LTA4	Leukotriene A4

LTA4H	Leukotriene A4 hydrolase
LTB4	Leukotriene B4
MAF	Minimum Allele Frequency
MDR	Multi-factor dimensionality reduction
MHC	Major histocompatibility complex
MZ	Monozygotic
NF-kB	Nuclear-factor Kappa B
NLR	Nod-like receptors
OR	Odds ratio
PB	Peripheral blood
PUFA	Polyunsaturated Fatty Acids
RR	Relative risk
SNP	Single-nucleotide polymorphism
TDT	Transmission disequilibrium test
TLR	Toll-like receptors
TNF	Tumor necrosis-factor
WTCCC	Wellcome Trust Case Control Consortium

To Maxine:

My brightest star in the darkest nights,

Je t'aime!

Acknowledgments

First and foremost, I am indebted to all Crohn's disease patients and their families who participated in this study. Ultimately it is them to whom this work is dedicated.

I would like to express my gratitude to my supervisor, Dr. Devendra Amre and his team at Ste. Justine's Hospital. Dr Amre's expertise, understanding, and patience, added considerably to my graduate experience. Over the years, he provided me with invaluable direction, support and friendship, for which I am forever grateful.

I am grateful to the Drs. Ernie Seidman, Emile Levy and all the other gastroenterologists at the 3 study centers. Equally, I am indebted to the Drs. Maja Krajinovic, Daniel Sinnett, Yves Theoret and Damian Labuda for the help they provided but also for sharing with me their years of knowledge and experience. For their support, I am indebted to all my friends and colleagues who work in their laboratories.

In particular, I would like to acknowledge the assistance of Drs Lei Sun, Kenneth Morgan as well as Michael Epstein and Jason Moore in the interpretation of the results.

For guidance and support during my academic formation, I would like to express my special gratitude to the Drs. Isabel Fortier, Nicole Leduc, Mark Daniel, and Marie-Pierre Dube, as well as the amazing Mme France Pinsonnault from the Département de Médicine Sociale et Préventive, Université de Montréal.

During my first year of doctoral studies, my research work was generously supported through a grant from APOGEE – Net, a network that unites a group of visionary people on the quest to support policy making in genetics. I would like to take this opportunity to thank them all. Particularly Dr Ingeborg Blancquaert, for providing me with invaluable insight into the decision-making process as it applies to genetics and public health. It was through this network that I met Dr Bartha M. Knoppers, the holder of the Canada Research Chair in Medicine and Law. Dr Knoppers is one of the professors/teachers who truly made a difference in my life. She provided me with friendship, motivation, encouragement, direction and financial support.

In the following years, I received a PhD grant from the Fonds de la Recherche en Sante - Quebec (FRSQ). I gratefully acknowledge their support.

I would like also to express my gratitude to the members of my PhD committee for taking time out from their busy schedule to review this thesis.

Finally, my family: to my parents, a special thank you, partly for the genes they gave me, but foremost for my upbringing that made me the person I am today; and to my best friends and most important people in my life, my husband, Costea and my daughter, Max – Thank you for making this possible!

Section A. Introduction and Background

The chronic inflammatory bowel diseases (IBD) are now common causes of gastrointestinal illness in developed countries. Recent population-based studies of disease prevalence suggest that as many as 0.5% of the Canadian population may be affected by chronic IBD [1]. Although mortality associated with IBD is uncommon, the diseases continue to have a high morbidity, particularly in children, in whom growth, psychosocial wellbeing and implicitly, quality of life are adversely influenced.

A.1. Crohn's Disease (CD)

Crohn's disease (CD) is a type of IBD characterized by chronic, relapsing inflammation of the gastrointestinal tract (GIT). First termed "regional ileitis" due to its predilection for the terminal small bowel [2], it is now well known that CD can affect any part of the GIT – **Figure 1**[*].

A.1.1. Epidemiology of CD

Epidemiologic studies show that the incidence and prevalence of CD varies significantly depending on geographic location and racial or ethnic background. Over the past few decades, incidence of CD has increased in the Western populations, especially in North-Americans as well as in populations of Northern European and Anglo-Saxon ethnic derivation, and seems to be on the rise in populations of developing countries, such as African populations [3]. Within North-America, as well as within Europe, a North-to-South gradient in the distribution of CD has been observed [4, 5]. Other studies however do not support this north-to-south gradient [6]. Although still a rare disease outside Europe and North-America, the incidence of CD in developing countries is also on the rise, e.g. Asia [7]. CD occurs about equally in both sexes and is more common among Ashkenazi Jews [8].

Till recently, CD was considered an adult-age condition characterized by 2 peaks: one between 25-40 years of age and a second one after 55 years of age [9]. In the last

[*] Figures and tables are presented at the end of each section.

decade, however, it is becoming more and more apparent that CD incidence and prevalence among children and young adults is on the rise.

As for the social determinants of adult-onset CD, it has been shown that there is an increased risk for developing CD in urban compared to rural areas, in higher socioeconomic classes, and in populations migrating from low-risk to high-risk geographical areas. Interestingly, the observed geographical trends seem thus, to point at environmental factors probably associated with adoption of a Western lifestyle and improved socioeconomic circumstances [10]. However, in the case of early-onset of disease, socio-economic factors (such as parental education and family income) have not been shown to be associated with CD [11].

Two population-based studies (i.e. one provincial and one nation-wide) conducted by Bernstein et al. revealed that over the last two decades Canada has one of the highest CD incidence and prevalence rates, i.e. an estimated 13.4/100,000 incident cases across Canada and 234/100,000 prevalent cases (year 2000), equivalent to a total of 170,000 existing cases (or approximately 1 in 180) (0.5% of the Canadian population). In addition, regional, age- and gender-specific differences have been noted. While the first study, restricted to the province of Manitoba, estimated the overall incidence and point prevalence (year 1994) for CD at around $14.6/10^5$ and $198.5/10^5$, respectively; the nation-wide study reveals, in addition to a slight increase in these rates in Manitoba ($15.4/10^5$ and $271.4/10^5$ - year 2000, respectively), also a east-west gradient, with the highest rates recorded in Nova Scotia, $20.2/10^5$ and $318.5/10^5$ (year 2000), respectively (**Figures 2a,b**). These regional differences were maintained by sex, age group, and urban/rural residence. In both studies, there was a sharp peak in the incidence rate among those aged 20-29 years ($28.4/10^5$), followed by those aged 30-39 ($17.6/10^5$) (**Figure 2a**). The peak prevalence was noted in the age group 30-39 years (around $310/10^5$), which is one decade later than the peak incidence (**Figure 2b**). The overall incidence rate for CD for children less than 20 years of age was $8.3/10^5$. Overall, the incidence of CD was higher among females for all ages, except for those less than 20 years, than among males, by a factor of 20-30% with roughly equal distribution in the urban versus the rural areas.

In a pediatric population, similar incidence rates among girls (i.e. $8.0/10^5$) compared with boys (i.e. $7.5/10^5$) were noted; however, with a higher CD prevalence rate among boys (i.e. $49.6/10^5$) when compared with girls (i.e. $43.8/10^5$). The overall urban to rural incidence ratio, adjusted for age, gender, and province was 1.05 [1, 9].

Overall these findings reveal: (i) an increase of CD burden over the past two decades, and especially among children and teenagers, with incidence rates approaching that in adults; (ii) a geographic gradient; and, (iii) a female preponderance in the adult population versus a smaller male preponderance in the pediatric population.

A.1.2. Demographics of CD in children

As indicated above, CD incidence and prevalence among children and young adults appears to be on the rise. In addition to the Canadian studies, ages at diagnosis and gender of children with CD have been characterized in a number of recent population-based epidemiologic studies. A recent study notes that 4% of the pediatric cases of CD are presenting before the age of 5 [12]. Another population-based study from Denmark approximates the mean age at diagnosis at around 12 years (range 0-14 years), indicating a steep increase in incidence around puberty [13]. However, further evidence linking factors implicated in pubertal changes to inflammation and CD development is still lacking. Unlike the adult cases, recent studies conducted in the United States, Canada, and the UK and restricted to pediatric-onset CD, document a male-to-female preponderance [1, 14, 15]. Although currently there is no convincing evidence that gender plays a role in the etiology or clinical outcome in CD [16], this contrasting higher male-to-female incidence of CD in the pediatric age group with a reversal of incidence and prevalence of CD in adults is intriguing.

A.1.3. Clinical features of CD in children

Early-onset CD is of important concern, because it presents numerous challenges and represents a particular burden. In comparison to adult patients, pediatric CD patients are more often closer to disease initiation and earlier in the process of chronic intestinal inflammation. More importantly, during the disease course affected children experience growth failure, the severity of the disease is greater and the frequency of complications is higher vis-à-vis adults. In particular, children with CD are more susceptible to relapses, undergo more frequent surgeries, are at high risk for growth retardation and pubertal delay, and thus suffer significant psychological morbidity [17].

A.1.3.1. Localization

Overall CD is characterized by a pan-enteric inflammatory process. All the gut wall layers can be inflamed and strictures and/or fistulae or abscesses can develop. Recent reports have shown that patients with early-onset CD may have specific phenotypes that differ from adult onset CD, thus suggesting that the pathogenesis of pediatric CD and adult CD may differ [18]. The most consistent difference in phenotype in children versus adults is in disease distribution. These reports suggest that compared to adult CD, first-decade pediatric onset CD is associated with more colitis and less ileitis [19-23]. Of particular interest is the subset of patients with a very early onset IBD (onset before the age of 8 year). These patients often present with rectal bleeding and have more isolated colonic involvement [18, 21, 24]. In a large study of nearly 1,400 North American early-onset patients, Heyman et al. demonstrated that a colonic predominant phenotype exists in CD diagnosed under the age of 8 years [21]. Paul et al. studied 413 pediatric IBD patients and also demonstrated a greater tendency for very young patients to present with colonic disease [24]. Finally, a recent study by van Limbergen et al on 276 patients with childhood-onset CD also points out the colonic predisposition in very young children [18]. Children with CD also have less ileal involvement as compared to adult-onset CD. The reasons for these phenotype differences are presently unclear but may be a reflection of differing genetic susceptibilities. Evidence for the latter comes from investigations that have shown that in early onset CD, those children who do not possess the NOD2/CARD15 disease conferring variants are more likely to have colonic disease [25]. Certainly genotype-phenotype investigations need to be further explored.

A.1.3.2. Symptoms and signs

The clinical presentation of CD is largely dependent on disease location and can include diarrhea, abdominal pain, fever, clinical signs of bowel obstruction, as well as passage of blood or mucus or both. The constellation of abdominal pain, diarrhea, poor appetite, and weight loss constitutes the classic presentation of CD in any age group [26]. Certain features however are unique to pediatric CD, as compared with adult onset of disease [16]. For instance, one important characteristic in children with CD is growth failure that can precede the development of intestinal symptoms and dominate the clinical presentation [27]. Impairment of linear growth can be present at diagnosis in up to 10-40% of affected children with CD. Several interrelated factors may contribute to growth failure in CD and some of these factors may underlie important pathogenesis mechanisms, as shown by the fact that further growth development is best achieved through control of intestinal inflammation [28]. Another more recent hypothesis suggests that immunity dysregulation directly determined by gut inflammation may affect growth. In particular, the direct growth-inhibiting effects of pro-inflammatory cytokines (i.e. small signalling proteins which mediate and regulate immunity), which are released from the inflamed intestine are incriminated [29].

A.1.3.3. Disease course

The natural history of CD is characterized by remission and relapses and occurrence of complications such as fistulas, abscesses, and strictures, the pathological process oft requiring surgical intervention. Furthermore, disease behaviour often changes during disease course. At diagnosis most patients have inflammatory disease (ranging from 70-90%) that can progress to stricturing (17%) and/or penetrating disease (characterized by fistulas and/or abscesses) (13%) [30] during disease course.

A.1.3.4. Disease classification

Patients with CD may show great variability in their disease phenotype (i.e. phenotypic heterogeneity) and response to treatment. It has been hypothesized that the phenotypic differences may reflect underlying differences in immunological processes between the different clinical entities [31]. In order to standardize the classification of the varying clinical phenotypes, the Vienna classification [32], and a more recent and refined classification, the Montreal Classification [33], have been promulgated. In brief, the Vienna classification of CD considered age of onset, disease location, and disease behaviour as the predominant phenotypic elements. The Montreal revision of the Vienna classification has not changed the three predominant parameters of age at diagnosis, location, and behaviour, but refinements based on epidemiologic and genetic observations within each of these categories have been made – **Table I**[†].

[†] Figures and tables are presented at the end of each section.

Despite important advances in diagnosis and treatment in recent years, the underlying cause of the disease remains unclear. As a result, the available therapies are not curative and may pose a substantial risk of side effects. Furthermore, because of the increasing burden of disease, demographic and clinical specificities, limited knowledge on whether potential risk factors contributing to adult-onset CD similarly predisposes to early-onset CD, investigating the etiopathogenesis of pediatric onset of CD is a public health priority.

A.2. Genetic epidemiology of CD

It is generally accepted that CD results from a complex interplay between environmental and genetic risk factors. Observations that CD incidence has increased markedly over the past half century, argues in favor of recent, unidentified, environmental contributions. Equally important, regional differences, twin studies and the familial aggregations of CD strongly point towards genetic risks. Furthermore, recent data suggests now that not only susceptibility, but also disease location, behavior and response to therapy may have a strong genetic influence [34]. Given the complex nature of the disease, i.e. the presence of phenomena such as incomplete penetrance, phenocopies, and genetic heterogeneity, it is more than likely that multiple genes acting in combination with environmental factors would enhance susceptibility for CD. CD is thus considered a multifactorial disease, with genetic, immunologic and environmental factors playing important roles in the development of the disease. This also means that CD is less likely to be due to a major gene effect, but rather due to multiple genes contributing modest risks to disease development.

Observations such as familial aggregation patterns, concordance rates in twin pairs and sibling pairs as well as ethnic differences have provided the impetus to study the genetic contribution in CD. More so, recent studies suggest that genetic factors may be of more importance in early-onset CD vis-à-vis the adult form [35].

A.2.1. Familial aggregation and ethnic differences

In pediatric CD the greatest risk factor for developing the disease is having other affected family members [12]. This observation was illustrated for the first time in a cohort study by Orholm et al. who showed that when compared with the general population, the first-degree relatives of a CD patient had a 10-fold increase in the risk of having the same disease as the proband [36]. This translates into an estimate of 5.2% of first degree relatives of non-Jewish probands to develop CD during their life time [37]. In addition, it seems that up to 75% of multiply affected families with CD are concordant for disease type [12].

In terms of ethnicity, it appears that while the highest prevalence is undoubtedly associated with Jewish ancestry [4, 37, 38], Asian Americans, Americans of Hispanic background, and aboriginal North Americans are less likely to develop CD [39]. The latter is confirmed by a study that reveals that aboriginals residing in Manitoba, Canada, are less likely to develop CD [40]. Nonetheless, other studies of migrant populations suggest that ethnic and racial differences may be more related to lifestyle and environmental influences than true genetic differences. For instance, until recently, CD occurrence was thought to be rare in the Indian subcontinent. Once these populations become exposed to a different environment and lifestyle, such as moving to the United Kingdom or Singapore, South Asians and their offspring are at increased risk of CD relative to British Caucasians [41] or to ethnic Chinese, respectively [42].

A.2.2. Twin studies

Twin studies have been a powerful tool in the identification of the different contributions of genes and environment to CD pathogenesis. This is because twins are generally brought up in the same environment, so differences in disease incidence between mono- (MZ) and dizygotic (DZ) pairs are used to estimate the genetic contribution to a phenotype. Although the true rate of twin concordance could only be estimated by large population studies, the early literature contained many case series reports of CD concordance in MZ twin pairs. These reports however were probably subject to reporting bias, with concordant pairs more likely to be documented, reported and published and thus likely to overestimate the genetic contribution to CD. A summary of all these early CD case reports found 85% of pairs concordant for CD [43].

In Europe, Tysk et al. [44] published the first population survey of twins and concordance for CD in 1988. The cases were selected from the Swedish twin registry (i.e. about 25,000 pairs of twins) and matched with hospital inpatient records; and 80 twin pairs suffering from IBD were identified (18 MZ twin pairs in the CD disease group). The original cohort of twins were followed up 15 years later and three further cases of CD were diagnosed [45]. All of these 'new' cases had been symptomatic at the time of the original survey but did not fulfil the diagnostic criteria. In the MZ - CD concordant pairs the authors were able to demonstrate, using the Vienna classification 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. There was no significant concordance for disease behaviour or extent. A second population study was performed in Denmark with similar results to the Swedish study. Orholm et al studied 103 twin pairs—again from a national registry—using a postal questionnaire to identify affected twin pairs. Two-fifths of CD-concordant pairs were concordant for disease location and two-fifths were diagnosed within 2 years of each other. The authors were able to calculate relative risks in this study and accordingly, a MZ twin of a CD proband has a relative risk for developing CD of 667 [46]. The third large published study was undertaken in the UK by Thompson et al. and examined 143 twin pairs. In this study, patients from a CD support group—the National Association for Colitis and Crohn's Disease—were sent a questionnaire to identify CD twin pairs. Unlike the other two studies, the concordance rate for CD was much lower (i.e. 20%, possible due to the selective nature of the study) than that of the two population studies [47]. However, a later extension of this study population to a total of 249 twin pairs did show similar concordance rates to the two population studies [48].

The individual concordance rates for the twin pairs from all three studies are shown in **Table II.** The combined results of the three studies reveal a MZ twin CD concordance rate of approximately 50% and less than 10% among DZ twins [49].

In addition, closer examination of the phenotypic characteristics of the Swedish cohort provided some evidence that in CD even the disease phenotype may be genetically determined. Thus, in the monozygotic CD concordant pairs 7/9 were concordant for disease location, as determined by Vienna classification, and 6/9 were diagnosed within 2 years of each other. There was no significant concordance for disease behaviour, i.e. fistulizing or stricturing type, or extent of the disease [45].

Overall the markedly higher MZ twin concordance rates, as compared with the DZ twins, demonstrates that susceptibility to IBD, particularly CD, is determined to a significant degree by genetic factors. Nonetheless, these values also show that MZ concordance was incomplete as the majority of MZ twins did not both have CD.

This suggests that significant non-genetic factors, primarily environmental, are required to “trigger” CD in a genetically susceptible host.

A.2.3. Parent-of-origin effects in CD

Observed phenomena in CD patients, such as familial aggregation [12, 36], increased prevalence among certain ethnic groups (e.g. Ashkenazi Jews) [4, 37, 38], and high concordance rate among identical twins [49], are highly suggestive of a genetic contribution to this disease. However, no simple genetic models are able to account for the familial patterns of disease. In addition to mutant genes and a hazardous environment, an interesting hypothesis that may be of relevancy in understanding the etiopathogenesis of CD is based on a phenomenon called genomic imprinting. Genomic imprinting can be viewed as an epigenetic system[‡], in which the activity of a gene is reversibly modified (“silenced”) depending on the gender of the parent that transmits it [50]. This may lead further to unequal expression from the maternal and paternal alleles of a diploid locus. The mechanisms for imprinting are still incompletely defined, but they involve epigenetic modifications that are laid down in the parental germ cells and consist in reversible changes in DNA methylation and/or chromatin structure [51]. In addition to genetic and non-genetic studies, there is accumulating evidence that genomic imprinting (or parent-of-origin effects) may play a role in the ethiopathogenesis of many complex diseases. Specifically, the preferential transmission of disease from parents of one sex to children has been previously observed in a number of diseases, such as Prader-Willi/Angelman syndromes, Alzheimer disease, autism, bipolar disorder, diabetes, obesity; as well as a number of cancers: bladder, breast, cervical, colorectal, esophageal, hepatocellular, lung, mesothelioma, ovarian, prostate, testicular, and leukemia [52, 53]. While the epigenetic system that is causing the genetic imprinting is difficult to study, it is generally accepted that epigenetic phenomena are at play in the following instances: (i) there is evidence for

[‡] Epigenetics is the adaptation of chromosome regions so as to perpetuate local activity states, whether of long or short duration and whether inherited or not.

significant heritability, (ii) most genetic linkage and association studies are negative, (iii) there is evidence of environmental determinism, and (iv) there is evidence for change in incidence over time [54].

As it applies to CD, the parent-of-origin hypothesis has been put forward based on a number of clinical and molecular findings in CD studies such as: (i) the maternal effect in transmission of CD from affected parent to offspring; (ii) parental differences in the degree of genetic anticipation; and, (iii) discordance of monozygotic twins affected by CD. This evidence is suggestive that inherited and/or acquired epigenetic defects, or epimutations, may be of aetiological and pathogenic importance in CD [55].

Clinical evidence for a parent-of-origin effect was recently shown in the analysis of 135 families in which both a parent and an offspring had CD. In the majority of families the transmission of susceptibility to disease occurred from mother to child, and the difference between maternal versus paternal transmissions was highly significant (p=0.00001) [56]. Interestingly, the maternal effect was observed only among non-Jewish pairs with CD in which affected mother-offspring pairs comprised 85% of the sample. This maternal effect is suggestive of several possibilities, such as mitochondrial transmission, X linked effects, and genomic imprinting. However, in the context of CD – viz. delayed onset (i.e. post-neonatal), remissions and relapses of this disease, a defect in mitochondrial genome is not likely to be the main cause of the maternal effect. X linked factors are of interest as an association of IBD with Turner's syndrome (i.e. a X monosomy) has been identified [57, 58], and evidence of CD linkage to markers on the X chromosome was recently detected [59]. The third possibility is genomic imprinting, or, as it has been explained, the differential expression of homologous genes depending on their parental origin.

Parent-of-origin effects were also detected in genetic anticipation studies of CD. Genetic anticipation is a biological phenomenon that exhibits an earlier age of disease onset and increased severity in affected offspring in comparison with their affected parents [60]. Anticipation studies, however, are not straightforward because a number of ascertainment biases may occur, and it has been shown that the rate of false positives can be very high if inappropriate sample collection criteria are used [61, 62]. One way for differentiating between artefactual anticipation and a genuine effect uses the comparison of parental effects on the degree of anticipation [63]. The rationale behind this idea is that if anticipation occurs as a consequence of ascertainment bias, it would be equally apparent in father-offspring and mother-offspring pairs, while anticipation confined to parents of one sex indicates some specific genetic or epigenetic event. This approach was applied in a sample of 61 parent-offspring pairs, both with CD, and it was found that the degree of anticipation was significantly greater for father-child pairs [20.6 (3.2) years, n=17] than for mother-child pairs [11.7 (2.1) years, n=44] [64]. Parental effects were also detected for severity where disease was more severe in the younger generation when the affected parent was the father (13 of 16 pairs) and not the mother (two of 11 pairs) [65]. Several other studies of genetic anticipation were less conclusive, with the trend towards higher degrees of anticipation in transmitting fathers [66] and transmitting mothers [67]. Discrepancies between the studies are not surprising given the differences in ascertainment criteria, ethnic background of CD, and the high degree of heterogeneity of CD.

As mentioned previously, twin studies detected a significantly higher concordance rate among MZ twins for CD compared with DZ twins [47]. Nonetheless, the combined concordance rate for monozygotic twins was relatively low (less than 50%) [49]. It is thought that environmental factors may cause phenotypic discordance of MZ twins as these twins are considered to be genetically identical. However, it would be very difficult to assess the large variety of environmental factors that may play a role in the pathogenesis of CD for each specific twin affected with CD. In this situation, epigenetic developments may be of interest to study. To some extent the epigenetic status of a gene represents an interface between the DNA sequence and the intra- and extracellular environment. DNA modification is subject to change due to the influence of cellular-environmental factors, and therefore epigenetic status represents a combination of both inherited and acquired traits. Thus, MZ twins, although carrying identical or very similar DNA sequences, may be very different from an epigenetic point of view [68].

Taken together, these findings suggest that parent-of-origin effects may be at play in CD. Exploring such effects in CD may further our understanding of the pathogenesis of this complex disease.

Considerable progress has been made in in the last decade in the identification and characterization of genes, as well as in understanding their functional role that implicates these genes in the etiology and pathophysiology of CD. Several approaches are presently used in the search for genetic susceptibility genes in CD: linkage studies, candidate gene association studies and genome-wide association studies (GWAS). In the next section we highlight the major findings based on these approaches.

A.2.4. Linkage studies

The principle of genetic linkage studies is to analyze families containing more than one affected member for the purposes of identifying chromosomal regions shared in excess of statistical expectation. A measurement of familial disease clustering, often used to estimate the number of families required to identify genes by linkage mapping, is *lambda*, the ratio of recurrence risk in siblings compared with population disease prevalence. For CD, a British report estimated *lambda* as high as 36.5; a high value when compared to other complex diseases (e.g. type II diabetes – 15) [69]. This observation suggests that CD may be controlled by relatively stronger genetic factors then other complex entities.

Of particular relevance are the genetic linkage studies that led to the identification of NOD2/CARD15 gene on chromosome 16q12 (IBD1) [70]. In brief, Hugot and colleagues conducted a linkage study of affected sibling pairs and identified an area of interest located in the peri-centromeric region of chromosome 16 (i.e. 16cen). Significant linkage to this chromosomic region, further designated as IBD1, translates into 62% of CD sibling pairs co-inheriting the same grandparental copy of IBD1 [71]. This also correlates with a relative sibling risk of 1.7 [72]. This degree of sharing was significantly greater than that expected, and therefore this finding was suggestive that the IBD1 region contained one or more CD predisposing genes. Further replication studies in Australia [73] and recommendations from the International IBD Genetics Consortium [71] made IBD1 a promising region for potential CD predisposing genes. This further led to the discovery, by both linkage disequilibrium fine mapping [70] and a focused candidate gene approach [74], of CD predisposing mutations within the NOD2/CARD15 gene located at chromosome 16, i.e. 16q12. The sequence of genetic discoveries from identifying an otherwise unknown CD predisposing locus, IBD1, to identifying NOD2/CARD15 CD predisposing mutations represented a major landmark for establishing an initial etiopathogenesis for an otherwise enigmatic immune mediated disease, as well as one of the first successes demonstrating the feasibility of positional identification of disease genes for a common complex genetic disorder. Identification and confirmation of the IBD1 locus has provided a strong impetus to search for other CD susceptibility loci by means of linkage analyses. In summary, genome wide linkage studies have now identified nine IBD designated loci (IBD1-9) on chromosomes 16q [72], 12 [72], 6p [59, 75], 14q [76, 77], 5q [75, 76, 78], 19 [75], 1p [79], 16p (overlap with IBD1)[80], and 3p [72] with specific identifiable CD risk factors [72, 77, 81]. In particular, risk alleles/haplotypes have been defined for the IBD1 (NOD2/CARD15) [70, 82], IBD3 (HLA) [83] and IBD5 (5q cytokine cluster) [78]. Several other regions show great promise for containing additional IBD loci, particularly chromosomes 2q, 3q, 4q, 7, 11p, and Xp each with suggestive evidence of linkage [84].

These findings are summarized in **Figure 3.**

A.2.5. Candidate gene studies in CD

Most genetic studies in CD have concentrated on a candidate-gene approach. In candidate-gene studies, the frequency of allelic variants of candidate genes – usually variants encoding proteins believed to be involved in pathogenesis – are compared between cases and matched controls (case control analysis), or by investigating intra-familial association/linkage (family-based studies). A number of candidate genes have been explored, but few have been validated and/or replicated. Of relevance to CD is the study by Ogura et al. who conducted a candidate gene approach and succeeded in revealing the role of NOD2/CARD15 gene in susceptibility to CD [74], at the very same time as Hugot et al. in France was publishing his linkage findings [70]. Research around NOD2/CARD15 gene is based on the assumption that the gene product of this gene, i.e. CARD15, is involved in the regulation of nuclear factor kappa-B (NF-kappaB) activation in response to bacterial endotoxins and hence may alter signalling pathways of the innate immune system [85]. Altered immune regulation is thus an important feature in the incriminated pathology of CD and genes thought to play a role in the regulation of innate immune response, mucosal integrity and apoptosis became subsequently primary candidates in the search for susceptibility genes in CD. Candidate gene studies have further led to the identification of several other susceptibility genes, including DLG5 [86], as well as novel organic cation transporter (OCTN) 1 and 2 [87], and NOD1 (CARD4) [88].

Overall, the candidate-gene approach has however proven to be difficult in confirming CD susceptibility genes. For instance, using a candidate gene approach, Peltekova et al. identified two genes, i.e. OCTN1/2 genes (i.e. IBD5 locus) that may confer susceptibility to IBD [87]. Nonetheless, due to tight linkage disequilibrium across the IBD5 locus and the wealth of candidate genes in the region [89], subsequently, it has been very difficult to identify with certainty the causative mutations in this region. Similarly, earlier studies of association with genes such as the DLG5 [86], PXR1 [90] and MDR1 [91] have not been consistently replicated in subsequent studies [92, 93]. A number of reasons could contribute to non-replication of candidate gene studies. Foremost is perhaps inadequate power and the tendency for early positive studies to be false-positive. On the other hand however, failure to replicate may be due to genetic heterogeneity, gene-environment interactions and phenotype heterogeneity. Some evidence for the latter comes from studies that have shown that the effects of the DLG5 gene may be different among males and females (please see below). Similarly, it is also more and more apparent that certain genes may affect specific clinical phenotypes of CD and lack of stratification for them may result in non-replication of earlier findings. This is in particularly apparent for early-onset CD (see below). For example we have recently shown that the MDR1 gene was not associated with overall risks for CD (*Krupoves et al, 2008, in press*), however associations were apparent with colonic CD and inflammatory disease, highlighting the need to investigate and assess lack of replication of candidate-gene studies.

A.2.6. Genome-wide association studies (GWAS) in CD

There is increasing evidence that GWAS represent a powerful approach to the identification of genes involved in complex genetic traits [94]. In essence, GWAS compare the frequency of alleles and genotypes between cases and related or unrelated controls on a genome-wide scale, thus creating a comprehensive unbiased method to identify candidate genes.

With respect to CD, access to GWAS has not only managed to identify new susceptibility loci [95-99] but more importantly implication of these loci is suggestive of novel mechanisms of disease susceptibility, as shown in a study conducted by the Wellcome Trust Case Control Consortium (WTCCC) [100-102]. As such, in one of the first GWAS, Duerr et al., by using the Illumina HumanHap300 Genotyping BeadChip (i.e. over 300,000 single nucleotide polymorphisms, SNPs) and a study population of non-Jewish, European ancestry patients with ileal CD and non-Jewish controls, reported strong association between variants in the IL23R receptor gene and CD susceptibility [95]. In addition to being highly significant and replicable, both in case-control and case-parent designs [95, 103], these data also provide insight to the importance of the IL23 pathway in CD pathogenesis.

To further advance gene discovery, Barrett et al recently combined data from the three genome-wide scans, i.e. The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) IBD Genetics Consortium [99], WTCCC [100] and the Belgian-French IBD Cohort Study [97] studies on CD (a total of 3,230 cases and 4,829 controls) and carried out replication in 3,664 independent cases with a mixture of population-based and family-based controls. The results strongly confirmed 11 previously reported loci and provide genome-wide significant evidence for 21 additional loci. Included among these are compelling functional candidates such as STAT3, JAK2 and IL12B, whereas others, such as CDKAL1 and PTPN22, highlight potentially intriguing contrasts between genetic susceptibility to CD and other complex disorders - viz. type I [104] and II [105, 106] diabetes and rheumatoid arthritis [107].

More recently, in the only GWAS focussed on the pediatric population, Kugathasan et al in a North American GWAS (647 CD and 4,250 matched controls) identified two previously unreported CD susceptibility (protective major allele was risk-conferring) loci (20q13 and 21q22, respectively). Although the exact genes could not been identified, based on the LD structure surrounding the identified SNPs, genes such as the TNFRSF6B and PSMG1 are considered to be implicated [108].

Findings from GWAS to date and as they relate to CD are summarized in **Table III**.

Although GWAS provide valuable clues to the pathogenesis of complex traits and there is growing focus on this approach, there are certain aspects related to GWAS for which considerable challenges remain. One such challenge relates in fact to the capacity of GWAS to provide a relatively unbiased examination of the entire genome for common risk variants; their weakness is that in doing so, these studies cannot clearly distinguish between the signal from true risk variants and the statistical noise from the vast numbers of markers that aren't associated with disease. To separate true signals from noise, researchers have to set a high threshold that a marker needs to exceed before it is accepted as a likely disease-causing candidate. That reduces the problem of false positives, but it also means that any true disease markers with small effects are lost in the background noise. Furthermore, most efforts at replication have concentrated on the signals for which the statistical evidence is strongest; whereas, some susceptibility loci with modest effect sizes might also benefit from further exploration in the context of their biological plausibility. As it stands, by increasing the numbers of samples in their disease and control groups, researchers are trying to dial down the statistical noise from non-associated markers until even disease genes with small effects stand out above the crowd. As the cost of genotyping is decreasing steadily such an approach will become more and more feasible; however, the logistical challenge of collecting large numbers of carefully-ascertained patients will always be a serious obstacle. This is because most published GWAS feature case-control designs (**Table III**) and thus concerns inherent to this type of design, i.e. selection bias, misclassification bias, and population stratification[§], remain.

One other important aspect is related to the fact that current genome scan technology rely heavily on the "common disease, common variant" (CDCV) assumption, which states that the genetic risk for common disease is mostly attributable to a relatively small number of common genetic variants. However, some fraction of the genetic risk of common diseases will be the result of rare variants, and findings from GWAS in a variety

[§] Current GWAS have been dominated by subjects of Western European ancestry, and our understanding of genetic risk variants in non-European populations is almost non-existent of diseases have failed to provide unambiguous support for the CDCV hypothesis [109]. Whatever the proportion of variance that turns out to be explained by rare variants, current GWAS technologies have lower power to unravel it. Increasing sample sizes may be beneficial, but the inability of current chips to tag rare variation will still remain a problem [110]. The solution will be higher-density SNP chips incorporating lower frequency variants identified by large-scale sequencing projects; however, such approaches will have diminishing returns: as chip manufacturers lower the frequency of the variants on their chips, the number of probes that will have to be added to capture a reasonable fraction of total genetic variation will increase exponentially, with each new probe adding only a small increase in power.

In summary, all the above approaches have led to the identification of several susceptibility genes for CD, with NOD2/CARD15, IL23R and ATG16L1 being the most important. Furthermore, the resulting genetic knowledge has provided important insights into the pathophysiology of CD and is stimulating future research. In parallel, the researchers' attention also focused on genotype–phenotype association considering that at least in adults the genes identified so far via GWAS seems to be associated with ileal CD rather than other forms of CD. In the next section we highlight genotype-phenotype correlations relevant to CD and also describe findings pertinent to CD in children.

A.2.7. Genotype-phenotype correlations in CD

It is generally accepted that adults and children with CD do not share the same disease phenotype. A interesting difference that exists between early onset and late onset IBD is a male preponderance reported in paediatric CD [9, 14, 20], while female preponderance is only seen among patients diagnosed in adolescence (13–19 years) [111], in accordance with the overall higher incidence of CD in females. Pediatric onset CD also appears to be associated with less isolated ileitis compared with adult onset CD [19-23]. Furthermore, in comparison to cohorts of adult patients, at diagnosis children are more often closer to disease initiation and earlier in the process of chronic intestinal inflammation [16]. Taken together, it seems likely that paediatric CD represents a specific group of patients with particular phenotypic appearance (and perhaps also drug responsiveness) that might be driven by different mechanisms – in which different genes may be involved, and/or different host responses - in children versus adults, such as decreased ileal susceptibility or increased susceptibility to purely colonic and/or pan-enteric disease (ileum + colon + upper tract). The genotype-phenotype characteristics of CD that we are reviewing here include gender, location of disease, and disease behaviour.

A.2.7.1. Gender

Although men and women share most genetic information, they have dramatically different disease susceptibilities that go beyond the expected gender-specific diseases. Gender was shown to influence susceptibility to nearly all common diseases that affect both men and women, including atherosclerosis and diabetes and their preceding risk factors, e.g. hyperlipidemia, insulin resistance, inflammation [112]. It has been generally accepted that endocrinologic differences are involved in the sexual asymmetry of complex disease; however, specific molecular mechanisms of such hormonal effects have not been elucidated yet.

As it pertains to complex diseases, recent studies investigating key factors in lipid metabolism (e.g. apolipoprotein E – APOE) and obesity (e.g. perilipin – PLIN) revealed that the interplay between genes, gender, and environmental factors may modulate disease susceptibility. For example, in the Framingham Heart Study, complex interactions have been shown between a promoter polymorphism at the apolipoprotein A1 gene, gender and dietary polyunsaturated fatty acid intake that modulate plasma concentrations of high-density lipoprotein cholesterol [113]. The gender-APOE-disease interaction was also supported by another study testing the hypothesis that ischemic heart disease differs as a function of APOE genotype in women and men [114]. Of particular relevance are also the gender-dependent associations found between common polymorphisms at the PLIN locus and obesity risk that has been replicated in several populations around the world [115, 116]. Gender effects in complex disease susceptibility have been reported also in auto-immune disorders; for instance, female specific linkage to chromosomes 2, 6 and 11 was observed in osteoarthritis [117].

In the particular context of CD, a better understanding of the significance of specific disease-associated genetic polymorphisms in the context of gender may be of critical importance. In 2002, Fisher et al decided to re-analyse their genome search data [59], with stratification of families by the sex of affected individuals, to identify potential excess allele sharing specific to males or females. To this end, the authors revealed that the major histocompatibility region on chromosome 6p, referred to as IBD3, showed evidence of male-specific linkage with a maximum LOD score of 5.9 in CD male-affected families. Regions on chromosomes 11, 14 and 18 also showed strong evidence of linkage in male-affected families but not in female-affected families. No evidence of sex-specific linkage was found in the IBD1 or IBD2 candidate regions of chromosomes 16 and 12 [118].

Of particular relevance are the findings by Stoll et al, who in 2004 described an association between CD and DLG5 variants. In particular, a certain variant, i.e. R30Q, was found to be associated with CD in a German family cohort. The authors replicated the R30Q association with CD in a European case-control cohort and found a non-significant over-transmission of the 30Q allele in patients with CD in a German/UK family cohort [86]. While Daly et al managed to replicate only partially these findings (one of two case-control cohorts and in a family cohort) [119], other subsequent studies failed to replicate the association of DLG5 30Q with CD [120-126]. Only recently, Friedrichs et al, by examining gender-stratified data from the case-control cohorts showing association in the initial two DLG5 studies [86, 119], was able to reveal that the DLG5 30Q allele was a risk factor for CD in men but not in women. The authors managed to replicate these findings by genotyping the R30Q mutation in two additional control samples, i.e. a sample of 190 male and 271 female participants from Germany and in a newborn sample of 301 male and 299 female infants from the United States (US). For all three samples (viz. combined control cohort, the German population sample and the US sample) the DLG 30Q allele frequency was significant lower in male than in female subjects [127]. Subsequently, other independent studies confirmed that DLG5 R30Q variant is a female-specific protective factor in pediatric onset CD [128, 129]. In our pediatric cohort, we have observed that the DLG5 R30Q variant was not associated with overall susceptibility for CD, nor were there gender differences in risk (*Amre et al, 2008, submitted*). On the other hand, however we observed that other variants in the DLG5 gene conferred risks for CD only among females.

Although gender differences are not uncommonly found in susceptibility to complex diseases and, as previously mentioned, it has been generally accepted that endocrinologic and lifestyle differences underlie this observation, the presence of the male-female allele frequency differences in the newborn sample may suggest that these differences can be due to prenatal processes rather than selection later in life. Additional studies are required to further elucidate gender-specific risks related to the DLG5 gene.

A.2.7.2. Location and disease behaviour

Defining the initial location of CD is essential for a number of reasons. In general, there is stability of anatomic location over the course of the disease [30]. Besides, certain disease locations, as suggested by Farmer et al [130] over 30 years ago and confirmed more recently by Oostenbrug et al [131], are associated with a particular clinical course. Patients with ileocolonic location and those with perianal involvement, for example, often suffer a more complicated course than those with disease localization in the terminal ileum or colon only. Furthermore, upper gastrointestinal location is less often complicated by fistulization [132]. Third, by considering perianal disease separate from internal fistulization, the Montreal Classification recognizes the independent courses of these 2 entities [133]. Treatment regimens, prognostication, clinical trials, and genotype–phenotype studies can all be confounded by the failure to recognize this distinction.

The location of disease in most studies prior to the adoption of the Vienna followed by the Montreal classification divided location into small bowel and/or colon. Both Vienna and Montreal classifications identifies the disease site as one of four locations: upper gastrointestinal, ileal, ileocolonic and colonic. Behaviour in both classifications is divided into non-stricturing, non-penetrating (inflammatory), structuring and penetrating [32, 33].

Variability in disease location is an important characteristic of CD. Incidentally, it has been suggested that CD is not a single disease, but a spectrum of diseases, with disease location and behaviour key element distinguishing individual cases. It is unclear why certain individuals would have predilection for disease in a certain location of the GIT or present with either inflammatory or non-inflammatory disease. It is however interesting that, several studies investigating the concordance rates and phenotypes in multiplex family studies demonstrated patterns of concordance for disease location and behaviour in CD pointing to inherited familial phenotypes [62, 134-136]. More recent studies that were mainly driven by the observation that disease location can be a strong predictor /determinant of disease progression [137] (i.e. small bowel involvement may predict early stricturing and the anoperineal location is showing more penetrating complications, respectively), have shown that interactions of susceptibility and modifying genes can also influence the specific features of disease location, behaviour, and complications. Studies investigating NOD2/CARD15 variation and disease localization have shown very consistent findings. For instance, six independent studies, each of more than 200 CD patients, have shown that the presence of NOD2/CARD15 mutations is associated with the ileal form of CD [34, 138-142] and stenosis [138, 140], although it is not clear whether this is a primary association [34] as ileal disease and stricture formation are not independent [143]. One possible explanation of the association between NOD2/CARD15 variation and ileal involvement rests with the overexpression of NOD2/CARD15 gene in the Paneth cells of the terminal ileum [144]. It is possible that partial or complete lack of protein function may lead to loss of the ability of Paneth cells to respond to bacterial components. With respect to the latter association, a meta-analysis of 42 study cohorts found a modest over-representation of the three common NOD2/CARD15 variants (viz. SNP 8, 12, 13) among CD patients with stricturing disease behaviour [145]. However, this observation needs to be viewed with caution, as for any given patient with CD the clinical phenotype regarding stenosis evolves over time, while disease location may change to a lesser extent in time [30, 137]. Thus, it is generally accepted that high-risk alleles may influence primarily the disease location and the location in itself can be a strong predictor /determinant of disease progression, i.e. small bowel involvement predicting early stricturing and anoperineal location showing more penetrating complications, respectively. Nonetheless, additional data is still required to demonstrate such assumptions.

As for other phenotypic associations, the strong evidence of linkage of markers at the IBD3 locus on chromosome 6p21 to CD provided the impetus to research on the major susceptibility genes located in this region. The evidence for linkage was shown to be strongest in the vicinity of the major histocompatibility complex (MHC), which, given its central role in the immune response, has long been considered an excellent candidate locus for CD. While an independent association of DRB1*0701 and Cw*0802 with CD was reported, the classic autoimmune haplotype A1-B8-DR3 was specifically associated with colonic disease [34]. More so, in Canadian cohorts, HLA DRB1*0103 was found to be associated with the colonic form of CD [146].

As it pertains to the perianal location of the disease, there is evidence of association with the IBD5 genotype [147] and another study showed significant association with the DLG rs2165047 mutation carriership in a pediatric population [148].

As previously mentioned, the most consistent difference in phenotype in children versus adults is in disease distribution; notably, a colonic predilection in early-onset CD. Since both the intestinal microbiota as well as the mucosal immune system are still under development in children, researchers speculate that the contributions of innate immune defects to the pathogenesis of CD are inversely related to the age of disease onset [149]. Based on this hypothesis that CD results from aberrant mucosal immune responses to environmental factors such as commensal bacteria and based on the relatively short exposure time to environmental factors in children, genetic mutations that either enhance the susceptibility for or protect from these pathogenic responses are thought to play an increasing role in early onset disease. However, evidence in support of this and further investigating candidate genes with mutations that are likely to present more frequently in childhood and lead to colitis-predominant CD is currently scarce, with two studies reporting on the frequency of association between previously identified genetic variants and pediatric CD [35, 148] and only one study identifying a candidate gene for the colitis-predominant disease of childhood. In the latter, Sun et al revealed that the cytosine insertion mutation 3020insC mutation in the NOD2/CARD15 gene was significantly more common in their pediatric population than in patients with CD (26% versus 11% of the alleles). The genotype-phenotype analysis further showed that the patients with at least one of the six CARD15 disease-associated mutations had a high risk of inflammation located in the terminal ileum and ascending colon. More so, in 10 of 19 patients with two mutations, intestinal resection surgery was necessary because of structuring [35]. These findings were complemented by those by De Ridder et al who reported a higher than expected by chance occurrence of two genetic variants, viz.

polymorphisms 3020insC in NOD2/CARD15 and SNP rs3792876 in SLC22A4/5, in patients with pediatric-onset CD than in patients with adult-onset CD. Polymorphisms 3020insC in CARD15 and SNP rs2165047 in DLG5 were also found to be associated with specific phenotypes in this pediatric-onset CD cohort [148]. Further supporting evidence is provided by Levine et al who described an association between a polymorphism in the TNF promoter and isolated colitis with an increased likelihood of pediatric onset disease in an Israeli cohort by comparing the prevalence of the polymorphism to an adult cohort [22]. The latter also published a recent study that investigated if the age-related colitis pattern noticed in younger children with CD is determined by genotype, by differences in colonic or ileal susceptibility, or due to an age-related variability in NOD2/CARD15 mutations. Although NOD2/CARD15 mutations were previously shown to be associated predominantly with ileal involvement [150], the authors found increased colonic susceptibility in patients with early-onset disease carrying these mutations. The most striking finding was a significantly higher prevalence of isolated colitis, but not colonic involvement, in NOD2/CARD15 wild-type patients, which presented most commonly in early childhood and declined with age. This trend for more colitis lasted until age 11. Together, these findings demonstrate yet again that pediatric onset CD may be characterized by different genes that predispose to early onset and isolated colitis [25]. Studies evaluating other candidate genes or loci in pediatric cohorts have failed to show a predominance of early age of onset or a colitis-predominant phenotype [123, 151].

In many of these studies, analyses of disease localization is complicated by variation in clinical classification schemes and by changes in disease behaviour with time [12], which renders the overall available data inconsistent. These observations highlight the need for independent replication in adequately powered studies, ideally using a different genotyping technology.

Although assigning disease phenotype may be difficult, this represents a critical step in determining the successful outcome of genetic association studies, as was shown in a group study that demonstrated that the erroneous assignation of disease phenotype may lead to a 40% loss of power in linkage studies [152].

A.2.8. Gene-gene interactions

An important goal of human genetics is to identify DNA sequence variations or polymorphisms in human genes that confer an increased risk to particular diseases. This is a difficult challenge for common, complex multifactorial diseases such as CD that are likely the result of interactions between multiple environmental and genetic factors [153-156]. In respect to the latter, evidence of interaction between established CD risks alleles has been found. For example, a complex pattern of gene-gene interaction between DLG5 and NOD2/CARD15 [86] as well as IBD5, ATG16L1, IL23R, and NOD2/CARD15 risk alleles [157] has been described. While progress in elucidating gene-gene and gene-environment interactions in the ethiopathogenesis of CD are clearly underway, such interactions are difficult to detect and characterize using traditional parametric statistical methods such as logistic regression because of the sparseness of the data in high dimensions. That is, when interactions among multiple variables are considered, there are many contingency table cells that have very few or no data points. This can lead to parameter estimates that have very large standard errors resulting in an increase in type I errors [158, 159]. In addition, detecting gene–gene (and/or gene environment) interactions using traditional procedures for fitting logistic regression models can be problematic leading to an increase in type II errors and a decrease in power. For example, forward selection is limited because interactions are only tested for those variables that have a statistically significant independent main effect. Those DNA sequence variations that have an interaction effect but not a main effect will be probably missed. With backward elimination, a complete model that includes all main effects and all interaction terms may require too many degrees of freedom. Stepwise procedures are more flexible than either forward selection or backward elimination but can also suffer from requiring too many degrees of freedom. In consequence, detecting interactions among variables is an increasingly recognized problem in human genetics [160]. To address concerns about inaccurate parameter estimates and low power for identifying interactions in relatively small sample sizes, a nonparametric and genetic model-free approach called multifactor dimensionality reduction or MDR that uses a data reduction strategy has been recently developed [161]. With MDR, multilocus genotypes are pooled into high risk and low risk groups, effectively reducing the dimensionality of the genotype predictors to one dimension. While evidence suggests that MDR is able to identify gene–gene interactions in the absence of any statistically significant independent main effects in simulated data [161, 162], in sporadic breast cancer [161], and in essential hypertension [160], no application of this tool in the study of CD has been yet reported.

While continued progress has been made in identifying potential genetic risk factors, the equally important role of mechanistic pathways that are involved in the pathogenesis of CD remains quite elusive.

A.3. Pathogenesis of CD

Recently, substantial advances in the understanding of the molecular pathogenesis of CD have been made owing to three related lines of investigation. First, the identification of NOD2/CARD15 gene [70, 82] followed by several additional susceptibility loci, e.g. IBD5, IL23R and ATG16L1 [78, 87, 95-97, 99, 163] point to the importance of the GIT epithelial/mucosal barrier function and innate and adaptive immunity in disease pathogenesis. Second, efforts directed towards the identification of environmental factors implicate commensal bacteria or their products, as drivers of dysregulated immunity and CD. Third, murine models have helped unravel the pathogenesis of CD. Taken together, this evidence implies that alternative pathways involved in the innate or adaptive immunity exert feedback regulation that result in gut inflammation and disease.

A.3.1. The role of microbes in the development of CD

Accumulating evidence suggests that the dynamic balance between microbes, particularly commensal flora, and host defensive responses at the mucosal frontier has a pivotal role in the initiation and pathogenesis of chronic CD. Evidence supporting this hypothesis is provided by the observed therapeutic benefits of antibiotic treatment in a subgroup of CD patients and recent findings suggesting that probiotics can ameliorate the symptomatology of CD [164, 165]. Furthermore, the enteric flora of CD patients has been found more commonly than in control patient groups to include strains of certain pathogens, such as E. coli that are able to adhere to the epithelium, and with low frequency effect epithelial invasion [166, 167]. The importance of the flora is more directly supported by studies in a variety of murine strains in which ‘spontaneous’ chronic colitis seems to be entirely dependent on the presence of a luminal flora. Thus, colitis is not observed when mice are maintained in a germ-free state, but rapidly emerges when they are reconstituted with bacteria that are considered normal constituents of luminal flora [168, 169]. Furthermore, in some instances, it has been possible to induce colitis in a susceptible murine strain with a single species of normal bacteria [170]. These studies provide compelling evidence that the nature of the host defenses, rather than the biological properties of a luminal bacterial species *per se*, may determine the functional outcome of that interaction.

Unfortunately, our understanding of the microbial flora itself is quite incomplete. Insights into the microbial–host interrelationships are hampered by both the limited knowledge of the diversity and complexity of the microbial flora and the limitation of available tools to delineate these characteristics.

A.3.2. The role of mucosal homeostasis in the development of CD

From comparative studies of germ-free and colonised animals, supported by *in vitro* data, intestinal flora clearly affects mucosal structure and function [171-173]. Barrier function is provided by anatomical features that physically impede penetration of macromolecules and intact bacteria. Several molecular features of the pre-epithelial barrier and structure of the epithelial tight junctions that comprise this physical barrier have been defined and it has become evident that these junctions are dynamically regulated in response to signalling molecules, in particular pro-inflammatory cytokines (i.e. TNF-alfa, IL-17, interferon-gamma, chemokines, etc) and the underlying immune cell network [174]. The importance of the epithelial barrier in disease predisposition is supported by the finding of abnormal intestinal permeability in some first-degree relatives with CD [175-178]. Expression analysis in human CD biopsies has demonstrated downregulation of junctional complexes, although the mechanisms involved are unknown [179]. A recent GWAS reported that 5p13.1, a CD locus contained within a 1.25-Mb gene desert, is associated with disease susceptibility and the associated alleles correlated with quantitative expression of the prostaglandin receptor EP4 (PTGER4) [97]. PTGER4 is expressed in intestinal epithelial cells and regulates epithelial barrier functions; interestingly, PTGER4 knockout mice are susceptible to chemical-induced colitis [180].

The epithelium is in constant communication with luminal flora and the underlying dense network of innate and adaptive immune cells. Production of chemokines by the epithelium in response to pathogenic infection alerts the host to a breach in barrier function and directs the immune response to the site of infection [181]. More specifically, intestinal epithelial cells express Toll-like receptors (TLRs)[**], Nod-like receptors (NLRs)[††], and other receptors for different chemokines and antibody-specific Fc [182-184]. In this context, epithelial-cell-specific NF-kB activation or suppression seems to be a nodal point in the suppression and/or recruitment of immune responses in CD. Interestingly, some bacteria seem to be able to shortcut this process to minimize epithelial cell NF-kB activation [185]. Taken together, these observations highlight the importance of NF-kB signalling networks within the intestinal epithelium in sustaining normal mucosal homeostasis and in mediating pathogen-specific responses.

In addition to the epithelial barrier, other specialized cells are interspersed along the crypt villus axis to enhance protection against microbes and promote repair. For example, Paneth cells reside in the base of the crypt where they secrete antimicrobial peptides, including the alfa-defensins. Some observations suggest that a reduction in Paneth cell alfa-defensins may contribute to the pathogenesis of terminal ileal CD in patients with mutant NOD2/CARD15 [186, 187].

A.3.3. The role of the mucosal integrity in the development of CD

The mucosal immune system is primed to detect bacteria and antigens at the mucosal surface and to drive an appropriate response. The response must be nuanced

[**] Toll-like receptors (TLRs) are a class of single membrane-spanning non-catalytic receptors that recognize structurally conserved molecules derived from microbes once they have breached physical barriers such as the intestinal tract mucosa, and activate immune cell responses. They are believed to play a key role in the innate immune system.

[††] The NOD-like receptors (NLRs) are cytoplasmic proteins that may have a variety of functions in regulation of inflammatory and apoptotic responses. Approximately 20 of these proteins have been found in the mammalian genome and include a subfamily called the NODs. Current understanding suggests some of these proteins recognize endogenous or microbial molecules or stress responses and form oligomers that activate inflammatory caspases causing cleavage and activation of important inflammatory cytokines such as IL-1, and/or activate the NF-kB signaling pathway to induce production of inflammatory molecules.

between ‘tolerant’ and ‘active’ to distinguish between an innocuous commensal and pathogens, which can invade the epithelium and beyond. A number of sentinel cell populations in the intestinal mucosa continuously monitor luminal microbes. These include specialized microfold cells (M cells) that have the unique ability to sample and transport antigen from the GIT lumen (i.e. microbial products), which are further picked up by antigen-presenting cells. Peyer’s patches, isolated lymphoid follicles and the lamina propria work as inductive sites for the mucosal response, whereas the lamina propria functions as a recognition/effector site [188, 189]. Among the subsets of antigen-presenting cells, myeloid-derived dendritic cells are the dominant subtype in the intestinal lamina propria and show considerable functional plasticity depending on the location, state of maturation, and stage of inflammation. In response to TLR ligands, the immature dendritic cells produce cytokines, such as IL-23, which contributes to development of intestinal inflammation in murine models of colitis [190, 191]. Recent studies also implicate IL-4, B-lymphocytes and enteric flora in dendritic-cell-mediated granuloma formation in states of chronic intestinal inflammation [192]. Consistent with a baseline state of hyporesponsiveness, intestinal macrophages show attenuated proliferation and chemotactic activity in response to either microbial ligands or host cytokines/chemokines despite possessing the molecular mechanisms to elaborate strong phagocytic and bactericidal responses [193]. Following an inflammatory signal, circulating macrophages migrate to the intestinal mucosa and these cells, unlike resident macrophages, express NLRs and TLRs capable of rapid response and functional chemotactic receptors. Direct support for the role of macrophages in the regulation of CD has been obtained from analysis of mice with selective transcription disruption in their macrophages [194]. The tumor necrosis-factor (TNF) produced by non-lymphoid cells, mostly macrophages, was found to be essential for the development of colitis using the adoptive T-cell model of colitis induction. Recent murine studies have also shown that depletion of macrophages prevents development of colitis, which otherwise occurs owing to unregulated production of IL-12 and IL-23 by macrophages [195].

Altogether, it seems clear that the innate immune system, considered being the first line of immune defence, evolved as a means to monitor the microbial environment and to limit infection by invasive organisms. Distinct classes of receptors, broadly expressed in many cell types, and that recognize microbial molecular patterns are central to innate immunity. To date, several TLRs and NLRs have been identified and characterized. Furthermore, mutations in both TLRs and NLRs have been found to be associated with CD, suggesting that each detection system is key for regulating mucosal homeostasis [196-199].

Perhaps the most compelling evidence for the importance of microbial-mucosal interaction in the pathogenesis of CD was provided by the identification of NOD2/CARD15 as IBD1. The emerging genetic data have inspired intensive studies of NOD2/CARD15 biologic functions and these analyses have already substantively enhanced understanding of the signaling pathways coupling bacterial pathogens to the earliest phases of the host immune response [200, 201]. Expressed in Paneth cells, enterocytes, monocytes, macrophages, T cells, and certain subsets of dendritic cells, the NLR family member protein NOD2/CARD15 is an intracellular pattern-recognition motif that promotes host resistance towards bacterial muramyl dipeptide (N-acetylmuramyl-L-alanyl-D-isoglutamine, also referred as MDP) [202].

The mechanism by which NOD2-mediated functions contribute to intestinal immune homeostasis and how dysregulation of these functions in individuals with disease-associated NOD2/CARD15 polymorphisms contribute to the increased propensity to develop CD remain incompletely understood. However, collectively studies suggest that NOD2/CARD15 primarily functions in antibacterial immunity and that persistent bacterial survival might be a driver of persistent inflammatory responses in CD [144, 203, 204]. However, a more comprehensive understanding of the relationship between NOD2-dependent pathways and cellular processes for handling internalized bacteria is needed.

A.3.4. The role of the autophagy mechanism in CD development

As noted above, a working model of CD pathogenesis has evolved from a central focus on mucosal integrity to innate immunity. A constellation of findings, particularly in the last year, suggest a more focused orientation of that model on intracellular responses to low-level invasive bacteria. These include recent findings implicating alterations in autophagy and phagosomal function. Autophagy is a fundamental biological process defined as a cytoplasmic homeostasis pathway whereby cytoplasmic portions become sequestered by a membrane for delivery to lysosomes. In general, the autophagy pathway plays a part in protecting mammalian cells against various bacterial pathogens and the cytotoxic effect of bacterial toxins [205] represent a primary attempt to re-establish homeostasis and when autophagic capacity is overwhelmed apoptosis could be triggered. Bridging the innate and adaptive arms of the immune system, autophagy is also linked to adaptive immunity, both by the delivery of ligands via TLRs and NLRs to promote inflammation and by delivery of cytoplasmic antigen to the human leucocyte antigen (HLA) class II molecules for the cross-presentation necessary for immune recognition [206-208]. Recent studies have also demonstrated that autophagy plays an important role in clearance of apoptotic bodies [209]. Persistence of apoptotic bodies as a result of incomplete autophagy in complex tissues such as the intestinal mucosa in turn could contribute to persistent inflammation and autoimmunity seen in CD.

The relevance to CD is highlighted by the recent GWAS discovery that a synonymous SNP in the auto-phagocytic gene ATG16L1 is associated with increased risk for CD [96, 99]. These findings have been subsequently replicated in pediatric onset CD as well [93]. ATG16L1 is broadly expressed in the intestinal epithelium, M cells, CD4/8 T cells, B1 cells and memory B cells. In addition, preliminary data have implicated ATG16L1 in host responses to intracellular bacteria [99]. Subsequently these findings were confirmed by a second study, part of the WTCCC GWAS [210]. The latter also directed attention towards a second autophagic gene for CD susceptibility – the IRGM gene. Specifically, studies reveal that IRGM is required for mycobacterial immunity and may have an analogous role in the granulomatous response often observed in CD [211, 212]. Future studies will need to systematically determine the functional implications of disease variants in a cell- and tissue-specific context.

A.3.5. The bridge between innate and adaptive immunity

Although innate immune responses seem to be a prerequisite for the excessive activation of adaptive immunity, the latter is the more proximate driver of tissue damage that is manifest in CD patients. Adaptive responses are executed by a combination of resident and recruited cell populations. These comprise mucosal B cells producing immunoglobulins A and G, and a complex mixture of T cells that are dominated by a Th1 phenotype in the case of CD (as opposed to Th2 phenotype for UC). Th1 development is triggered by microbes that stimulate production of specific cytokines, such as interferon-gamma and IL-23 [213-215].

Another CD4 T-cell lineage (Th17) has been described that is characterized by the production of the cytokine IL-17 and the development of which is promoted by IL-23 and suppressed by transcription factors required for both Th1 and Th2 cells [215, 216]. Although the precise mechanism by which IL-23 maintains Th17 responses *in vivo* is still not well understood, recent studies have shown that Th17 cell lineage commitment is driven by transforming growth factor-beta (TGF-beta) in the presence of pro-inflammatory cytokines, whereas IL-23 seems only to be able to expand or maintain Th17 cell populations [217]. Recent studies have demonstrated that an IL23R coding variant is associated with reduced risk of CD [93, 95]. Further analysis suggests that there are multiple variants in the region independently associated with risk of CD and establish that IL23R signalling is central to the pathogenesis of the disease. In addition to its ability to support the development of Th17 cells, IL-23 induces the secretion of IL-17 by non-T-cells in an inflammatory environment, and both T cells and monocytes serve as sources of increased expression in the mucosa of CD patients [174]. Recent data also suggest that IL-17 induces antimicrobial peptides and may regulate tight junction barrier formation [218]. Taken together, these data suggest that the IL-23 signalling pathway seems to function as a key conductor of innate and adaptive inflammatory responses in the intestinal mucosa [174].

An important conceptual development in the understanding of CD pathogenesis has been the more focused appreciation of the nature of the microbial–innate-immune-response interaction during the transition from physiological to pathological intestinal inflammation. As highlighted above, achieving a thorough understanding may depend as much on more rigorous analysis of the complex dynamics of luminal microbial communities as on understanding of the host mucosal defence and response mechanisms.

However, independent of the putative mechanism of disease, a characteristic feature of CD is uncontrolled and chronic inflammation in the GIT. Based on this fact, new formulations that target and inhibit leukotriene synthesis have been developed. The first anti-leukotriene drug, zileuton, has recently been evaluated in CD and a significant, although insufficient, clinical response was obtained by a 70% inhibition of rectal LTB4 synthesis [219]. Thus studying intrinsic pathways that lead to inflammation is likely to provide valuable insights into the pathogenesis of CD.

The proposed mechanisms underlying the pathophysiology of CD are summarized in **Figure 4**.

A.3.6. Inflammation and CD: mechanisms

As reviewed above, many factors contribute to the complex course of inflammatory reactions. Microbiological, immunological and toxic agents can initiate the inflammatory response by activating a variety of humoral and cellular mediators. In the early phase of inflammation, excessive amounts of interleukins and lipid-mediators are released and play a crucial role in the pathogenesis of organ dysfunction. A widely studied pathway of inflammation is that mediated by eicosanoids, which are pro-inflammatory signalling molecules.

A.3.6.1. The role of eicosanoids in the pathogenesis of inflammation and CD

Arachidonic acid (AA), the mother substance of the pro-inflammatory eicosanoids, is a normal constituent of cell-membrane phospholipids. Individuals living on a typical western diet have a large amount of AA acid in the cells. In particular, about 25% of the fatty acid in platelets is constituted by AA, whereas the proportions in mononuclear cells (22%), neutrophils (15%), and erythrocytes (17%) are also high [101]. Cell-bound AA is converted to free AA by the action of phospholipase A_2. Once it is released from membrane phospholipids in the course of inflammatory activation; the enzymes cyclooxygenase and peroxidase lead to prostaglandin H_2, which in turn is used to produce the prostaglandins, prostacyclin, and thromboxanes; and the enzyme 5-lipoxygenase (5-LO) leads to the production of leukotrienes (e.g. LTB_4) via the 5-lipoxygenase (5-LO) pathway [220] (**Figure 5**).

The end-products of AA metabolism are dependent on the specific cell, and the activity, expression and amounts of these enzymes. The production of these derivatives and their action in the body are collectively known as the “arachidonic acid cascade”.

Eicosanoids are thought to be implicated in the pathophysiology of inflammation because of their potent effects, in particular the effect of leukotriene B4 (LTB_4), on vascular tone and permeability, mucus secretion, and leukocyte recruitment [221].The 5-LO pathway by which LTB_4 comes about becomes activated by various stimuli (i.e. dietary factors, immune complexes, bacterial peptides) that elicit a sequence of events starting with the translocation of arachidonate 5-lipoxygenase (ALOX5), a key enzyme, to the nuclear envelope. Here, ALOX5 transforms the released AA to the epoxide leukotriene A4 (LTA_4), which can further undergo transformation by one or more of three possible fates depending on the cellular context: hydrolysis, conjugation with glutathione, or transcellular metabolism to generate bioactive eicosanoids [222]. In particular, the hydrolytic attack of LTA_4 in pro-inflammatory cells by leukotriene A_4 hydrolase (LTA_4H) enzyme present in the cytoplasm, and potentially in the nucleus, yields LTB_4, a inducer of inflammation and a powerful mediator of leukocyte chemotaxis and adherence [220, 223]. Unknown transporters than facilitate efflux of LTB_4 out of the cells. LTB_4 can than act on target cells that can include epithelial cells, smooth muscle cells or endothelial cells by interacting with one of 2 receptors: LT-B1 and LT-B2. LT-B1 is expressed primarily on leucocytes and is a high-affinity receptor whereas LT-B2 is expressed ubiquitously, but has a somewhat lower affinity for LTB_4. Activation of the receptors leads to the activation of the Gq class of G proteins that subsequently result in a myriad of downstream effects. These include accumulation of virtually all subgroups of leukocytes to the site of inflammation, increasing the expression of adhesion molecules (thereby enhancing leukocyte adhesion to the microvasculature), and promoting cell motility that leads to transmigration of cells to the tissues. These cascades of events lead to the characteristic process of inflammation.

The important role of LTB4 in the development and/or perpetuation of inflammation in CD is supported by several lines of evidence. In one of the earliest studies, Sharon et al (1984) examined the expression of LTB_4 in CD patients. In the colonic mucosa of the patients they analyzed lipid extracts using high pressure liquid chromatography. They observed that the mucosa from CD patients contained 254 ng/g of LTB_4 whereas that from normal subjects contained a significantly lower amount (5ng/g) [224]. In a subsequent study these authors observed that the chemotactic response in CD mucosa was blocked by anti- LTB_4 sera and furthermore that the amount of chemotactic activity correlated with the high levels of LTB_4 in the mucosa [225]. These observations were later confirmed by Lauritsen et al (1988) who demonstrated that LTB_4 was elevated in CD patients in particular in those who had rectal ulcerations [226]. More recently, Schmidt et al (1995) also demonstrated higher levels of LTB_4 in colonic mucosa of untreated and active CD patients [227]. In spite of the reported observations of elevated levels of LTB_4 in IBD and CD patients, few studies have examined the expression of the enzymes/proteins involved in LTB_4 synthesis in IBD. In one study, Hendel et al (2002) examined the mRNA expression of 5-LO in 21 patients with CD and 12 healthy controls. Using real-time polymerase chain reaction (RT-PCR) methods they reported non-significant elevated trends in the levels of 5-LO mRNA expression in patients who had either active CD or quiescent CD as compared to controls [228]. More recently Jupp et al (2007), in a comprehensive examination of LTB_4 synthesis pathway enzymes using immunohistochemistry reported a 3-fold higher number of cells staining for 5-LO, and a 4-fold higher number of cells staining for LTA_4H in colonic biopsies of patients with active CD as compared to healthy controls. In the small subset of CD patients (n=7), a higher number of cells stained for the LTA_4H enzyme. They further examined whether frequencies of leukocyte counts in the biopsies were increased. They observed that there was an 18-fold higher mean count of neutrophils in active CD biopsies, a 4-fold increase in the mean number of eosinophils and a 3-fold increase in the mean number of macrophages. Consistent with these observations they reported strong correlations between neutrophils and cells expressing 5-LO and LTA_4H in biopsies from CD patients [229]. These observations highlight that an abnormal prevalence of enzymes that co-ordinate to synthesize LTB_4 may be intimately linked to colonic tissue injury and inflammation in CD. It could thus be postulated that genetic variation in the synthesis and activities of these enzymes/proteins could underlie inter-individual differences in the production of LTB_4 that could determine susceptibility for inflammation and CD.

Because of the important role of LTB_4 in inflammation, the AA metabolic pathway provides important target molecules for potential therapy of inflammatory disorders such as asthma, atherosclerosis, psoriases and CD. Clinical trials of CD patients using 5-LO inhibitors however, have been inconclusive [230-232]. The variability in clinical response may reflect variation in the numbers of inflammatory leukocytes expressing the 5-LO pathway in colonic mucosa, which in turn may be influenced by immunological and genetic factors [233]. The latter is supported by studies in particular in asthma where genetic variations in the ALOX5 enzyme were shown to determine variability in response to anti-leukotriene drugs [234-236].

A.3.6.2. Genetic variation in genes involved in the 5-LO pathway and their potential role in CD

As activation of LTB_4 can trigger a cascade of downstream events leading to the inflammatory process its regulation is important in downgrading these responses. Such modulation can occur by the cellular expression of the 5-LO pathway, and/or the levels and activity of the enzymes involved in the 5-LO pathway, all of which may be under genetic control. There is currently no information on associations between genes coding for molecules within the LTB_4 biosynthesis and/or regulation pathway and CD. However, there is evidence from studies among clinical phenotypes that have a significant inflammatory component suggesting that genetic variation in genes coding for 5-LO and CYP4F2 could be important. In the following section we briefly outline this literature.

The ALOX5 gene that codes for 5-LO enzyme is situated on the long arm of chromosome 10 (10q11.2). The encoded protein, which is expressed specifically in bone marrow-derived cells, catalyzes the conversion of AA to LTA_4. Mutations in the promoter region of this gene have been shown to affect mRNA associated levels, which can be further linked to increased mitochondrial damage and apoptosis [237]. Such alterations may lead to inflammation in animal models and in humans [238, 239] and diminished response to antileukotriene drugs in the treatment of asthma (e.g. zileuton) [240, 241]. In a recent study, Dwyer et al (2004) studied carotid artery intima-media thickness in relation to the polymorphisms in the promoter region of the gene, specifically the number of tandem repeats. Of 6 alleles, the most frequent allele, accounting for 80.5%, contained 5 of these tandem motifs. Variation in the tandem repeats (i.e. addition or deletion) was shown to be associated with atherosclerosis. Of particular interest the same authors also reported diet-gene interactions suggesting that dietary-omega-6 polyunsaturated fatty acids (viz. AA precursors) promoted whereas omega-3 fatty acids inhibited leukotriene-mediated inflammation and lead to atherosclerosis in their population [242]. Similarly, there are suggestions that variants in particular the promoter region of the gene are associated with asthma or airway hyper-responsiveness [241, 243, 244] indicating that the gene may be related to the inflammation characteristic of these diseases.

At the level of the 5-LO pathway, other genes that can potentially affect inflammation are the genes that regulate the production of LTB_4. LTB_4 and some related eicosanoids are catalyzed/de-activated by the cytochrome P450 omega-hydroxylases (CYP-450) belonging to the CYP4F subfamilies, in particular the CYP4F2 enzyme. The CYP4F2 enzyme is an efficient LTB4 omega-hydroxylase, judging from its localization in human white blood cells and human liver and its substrate specificity, the Km[‡‡] value for LTB4. Determination of the Km value allows comparing the specificity of different substrates for the same enzyme or the comparison of hydrolysis rates with different enzymes cleaving the same substrate. For a given substrate, a low Km value, as in the case of CYP4F2, is illustrative of high specificity [245]. The role of this enzyme is to metabolize LTB_4 by omega-oxidation to 20-OH LTB_4 and 20-COOH LTB_4, essentially inactive products of LTB_4 [246]. Of importance are also observations that CYP4F2 is involved in the omega-hydroxylation of AA thus potentially regulating the biosynthesis of LTB_4 as well. Furthermore, studies have shown that it is also expressed in the intestine [247], thus signifying its potential role in deactivating both circulating as well as tissue-specific LTB_4.

Thus, it can be postulated that the CYP4F2 enzyme is a key regulator of the negative-feedback mechanism that influences the actions of LTB_4 in various tissues. The CYP4F2 enzyme is encoded by the CYP4F2 gene (NC_000019.8), part of a cluster of cytochrome P450 genes on chromosome 19 (i.e. 19pter-p13.11).

[‡‡] Km = Specificity constant

With regards to CYP4F2, three genome scans have revealed significant linkages to the chromosomal region where it is located (i.e. 19p) in different populations. Rioux et al (2000) using affected sib-pairs reported peak LOD scores[§§] of 4.6 for IBD and 3.0 for CD at 19p in a Canadian population [75]. In their recent meta-analyses of 10 IBD genome scans among affected relatives, van Heel et al (2004) also reported significant linkage of CD to chromosome 19p [248]. Finally, Low et al (2004) carried out a linkage scan among UK Caucasians (affected sib-pairs) and confirmed linkage of CD to the 19p13.2 region (peak multi-point linkage score of 1.59) [249]. Further support for the role of CYP4F2 in inflammation was provided by the study of Curley et al (2006). Based on the observation that celiac disease (i.e. another inflammatory disorder of the gastrointestinal system) shares two linkage regions with IBD, viz. on chromosome 5q31 (CELIAC2 and IBD5) and 19p13 (CELIAC4 and IBD6) [75, 248, 250, 251], the authors hypothesized that genes situated on overlapping regions of chromosomes 5q31 and 19p13.1 might be associated with susceptibility for both disorders, most probably by influencing inflammation in the gut. Further based on its role in inflammation and its location in the overlapping regions, CYP4F2 gene was thus chosen as one of the studied candidate genes. Two out of eight CYP4F2 tag-SNPs (selected according to their block-tagging ability, replication status, allele frequency and SNP density) showed positive association with celiac disease in this unrelated case-control study, further supporting a putative role for this gene in gut inflammation [252].

Evidence for linkage to IBD on chromosome 19p was also used as an argument by Tello-Ruiz et al to conduct a screen for candidate genes [253]. However, none of the CYP4 family genes were considered for further investigation.

Finally, a recent meta-analysis combining data from three GWA studies on CD (a total of 3,230 cases and 4,829 controls) and further replicated in 3,664 independent cases

[§§] LOD stands for logarithm of the odds (to the base 10). A LOD score of three or more is generally taken to indicate that two gene loci are close to each other on the chromosome. (A LOD score of three means the odds are a thousand to one in favor of genetic linkage).

with a mixture of population-based and family-based controls, identifies the 19p13 locus as a potential candidate region associated with CD [254].

In summary, LTB_4 is crucial to the inflammatory process. Its role has been well demonstrated in a number of inflammatory diseases. Both *in vitro* and *in vivo* studies show that LTB_4 expression is hampered in CD. Recent evidence highlights that enzymes/proteins in the LTB_4 synthesis pathway are abnormally expressed in CD mucosa and the frequency of cells expressing these proteins was substantially higher. Although limited, genetic studies suggest in particular that the CYP4F2 gene could be linked to CD. Furthermore the ALOX5 gene is associated with susceptibility for other inflammatory diseases suggesting that it may play important roles in CD as well.

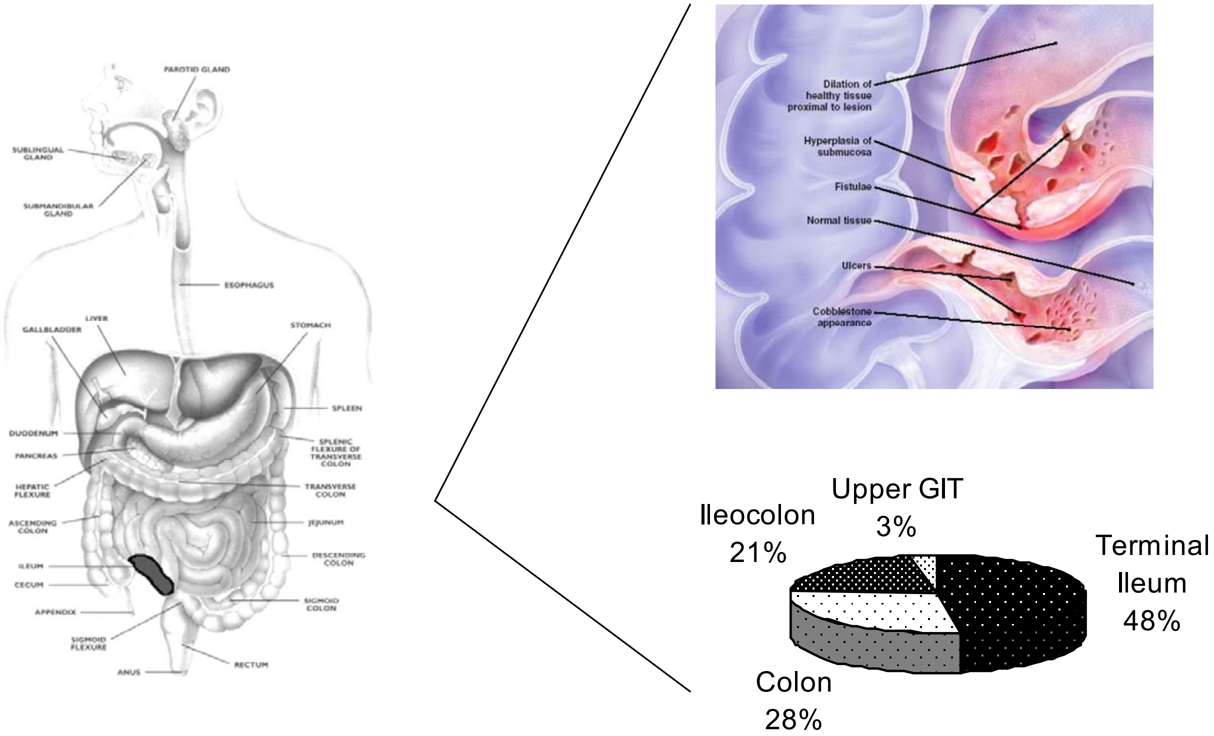


Figure 1 – Localisation and pathologic features of pediatric Crohn's disease

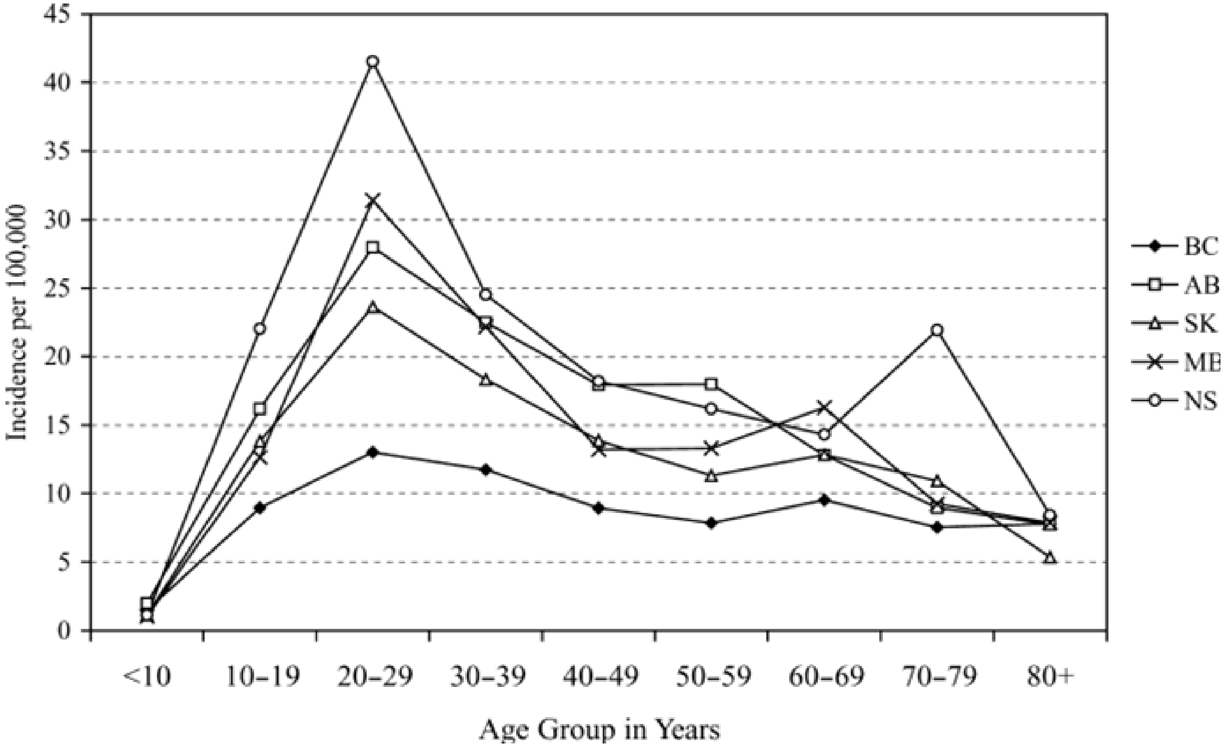


Figure 2A - Age-specific incidence per 100,000 of Crohn's disease (CD) in five Canadian provinces, 1998-2000

Legend: BC – British Columbia, AB – Alberta, SK – Saskatchewan, MB – Manitoba, NS – Nova Scotia

(from Bernstein, C.N. et al. « The Epidemiology of Inflammatory Bowel Disease in Canada: A Population-Based Study » *The American Journal of Gastroenterology* **101** (7), 1559-1568)

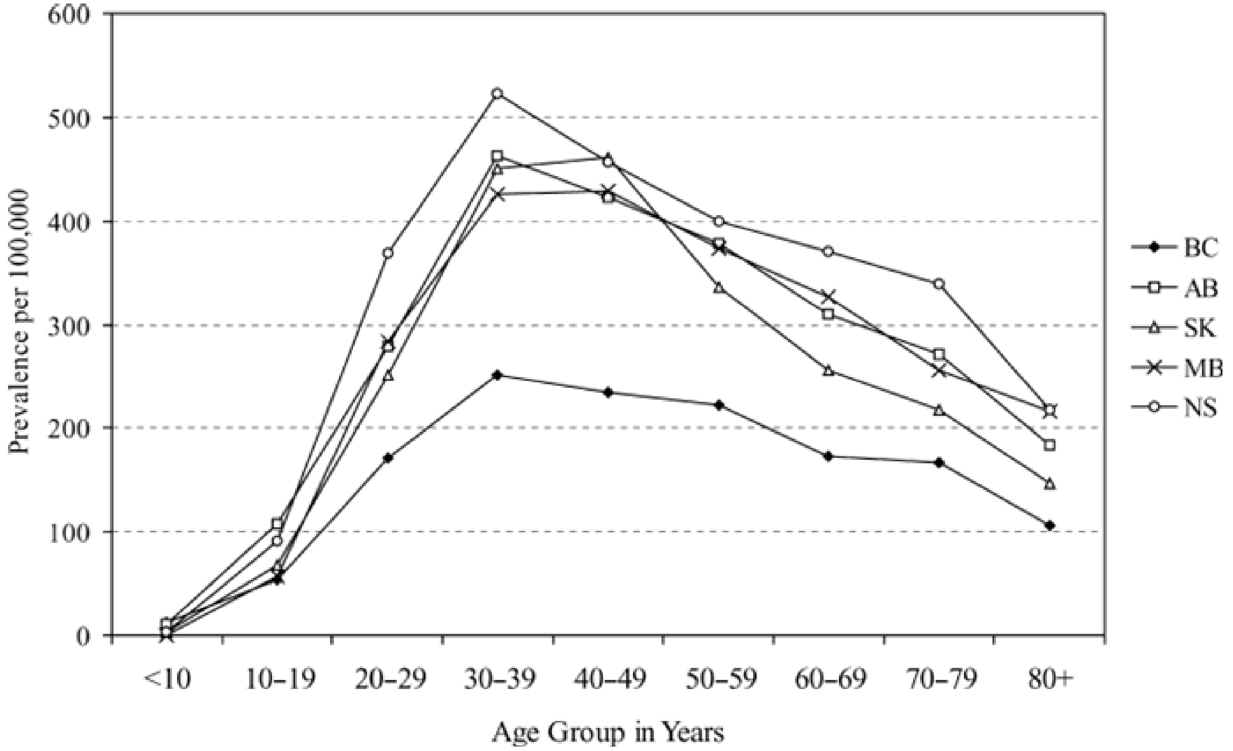


Figure 2B - Age-specific prevalence per 100,000 of Crohn's disease (CD) in five Canadian provinces, 1998-2000

Legend: BC – British Columbia, AB – Alberta, SK – Saskatchewan, MB – Manitoba, NS – Nova Scotia

(from Bernstein, C.N. et al. « The Epidemiology of Inflammatory Bowel Disease in Canada: A Population-Based Study » *The American Journal of Gastroenterology* **101** (7), 1559-1568)

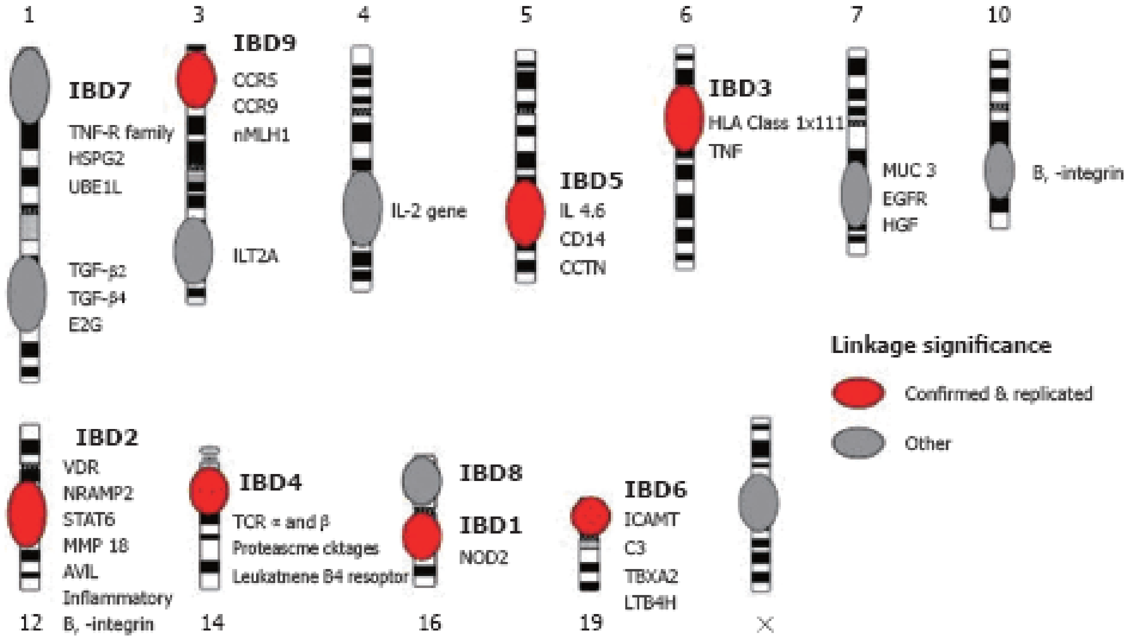


Figure 3 - Confirmed and suggested linkage intervals from genome wide scans in Crohn's disease (CD) with the potential candidate genes coloured in red (significance levels defined by Lander and Kruglyak)

(from Noble C et al. Novel susceptibility genes in inflammatory bowel disease. World J Gastroenterol. 2006 Apr 7;12(13):1991-9.)

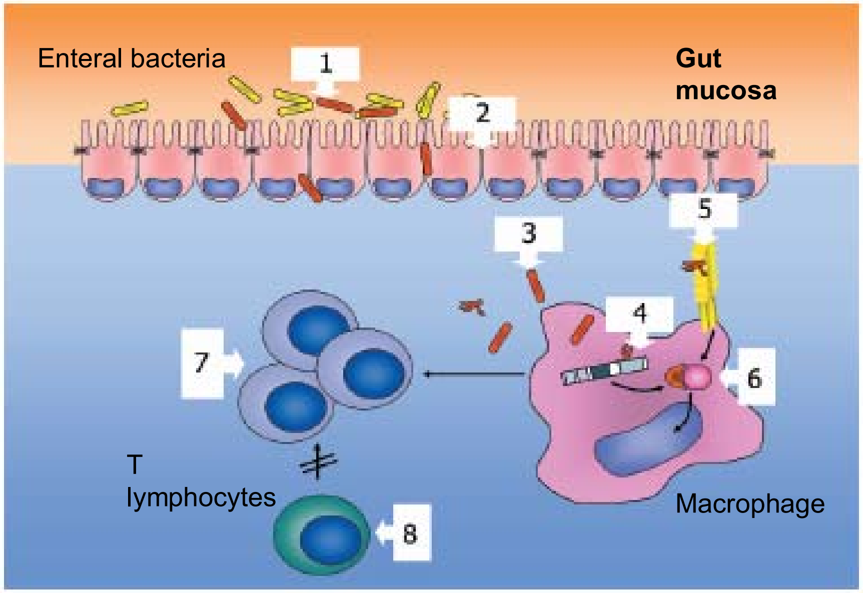


Figure 4 – Hypothesized mechanisms underlying the pathogenesis of Crohn's Disease (CD)

(1) A dysbalance of enteral bacteria changing epithelial barrier function or a primary barrier defect (2) allows bacteria an increased translocation across the epithelial barrier (3) challenging the mucosal immune system. Bacteria or bacterial wall compounds may enter macrophages/dendritic cells and bind to NOD2/CARD15 (4). In case of mutations with loss of function, no sufficient NF-kappa-B activation is achieved and the translocated bacteria cannot be quickly eliminated.

An insufficient response will also occur if bacterial sensors (i.e. toll-like receptors) in the cell membrane are mutated (5).

Another possibility is a mutation of NF-kappa-B itself (6) reducing protein expression of efficient translocation and inhibiting or delaying an immediate response. If the stimulus cannot be eliminated rapidly, antigen-presenting cells become activated and interact with Th1 cells. By expression of co-stimulatory molecules, which are normally downregulated, they induce T-cell proliferation (7) by secretion of Th1 stimulating cytokines such as IL-12, IL-23 or IL-18 to compensate the impaired innate immunity response. Insufficient activity of regulatory T-cells (8) could contribute to prolonged Th1 activity and chronicization of the inflammation.

(from Rogler, G. Update in inflammatory bowel disease pathogenesis. **Current Opinion in Gastroenterology** 2004, 20:311–317)

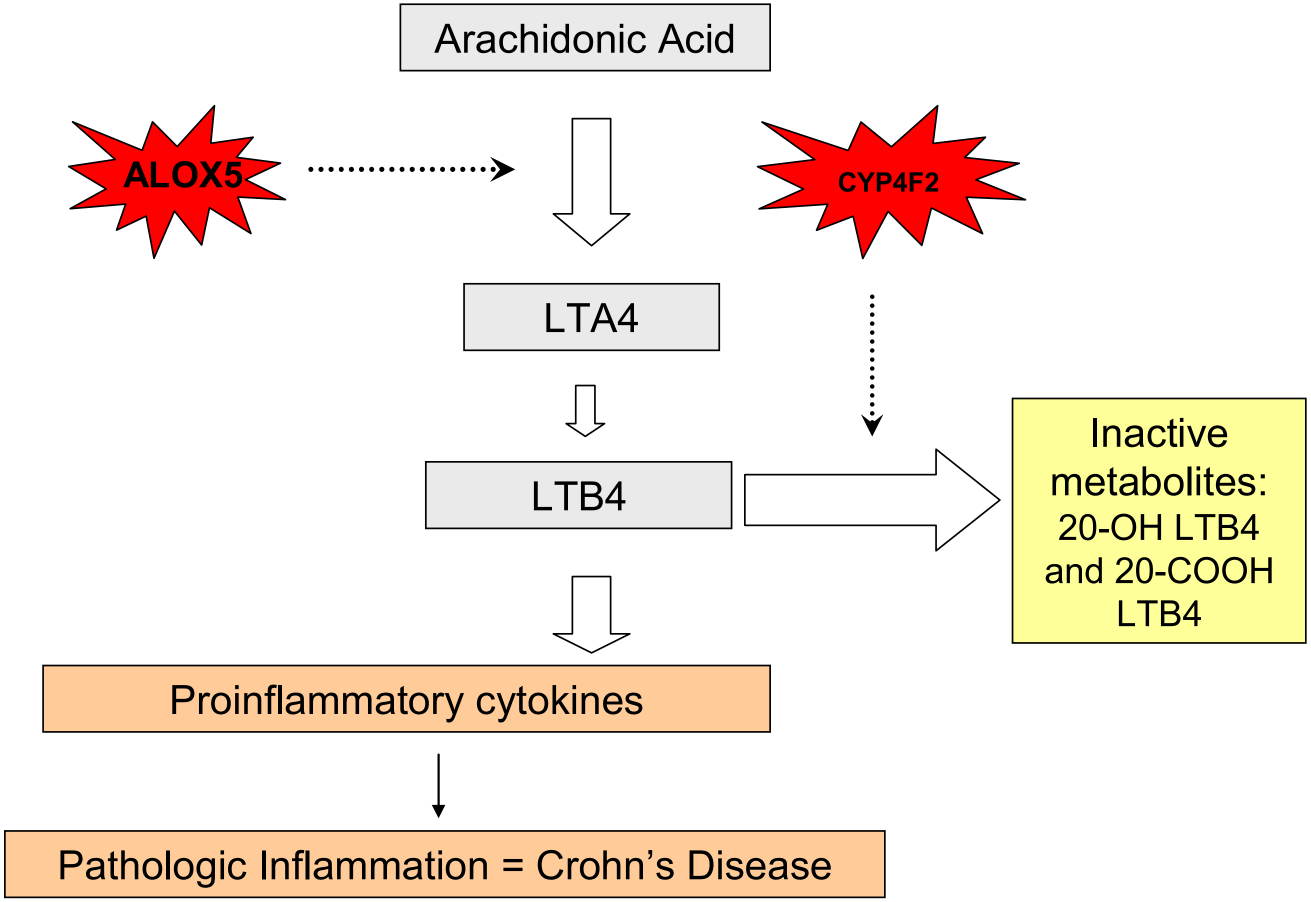

Arachidonic Acid
ALOX5
CYP4F2
LTA4
LTB4
Inactive metabolites: 20-OH LTB4 and 20-COOH LTB4
Proinflammatory cytokines
Pathologic Inflammation = Crohn's Disease

FIGURE 5 - 5—LIPOXYGENASE (5-LO) PATHWAY

Table I — crohn's disease (CD) Vienna and Montreal Classification

	Vienna Classification	**Montreal Classification**
Reference	(Gasche, Scholmerich et al. 2000)	(Silverberg, Satsangi et al. 2005)
Age at diagnosis (A)	A1 below 40 y	A1 below 16 y
	A2 above 40 y	A2 between 17 and 40 y
		A3 above 40 y
Location (L)	L1 ileal	L1 ileal
	L2 colonic	L2 colonic
	L3 ileocolonic	L3 ileocolonic
	L4 upper	L4 isolated upper disease*
Behavior (B)	B1 non-stricturing, non-penetrating	B1 non-stricturing, non-penetrating
	B2 stricturing	B2 stricturing
	B3 penetrating	B3 penetrating
		** « p » perianal disease modifier

*L4 is a modifier that can be added to L1–L3 when concomitant upper gastrointestinal disease is present.

** "p" is added to B1–B3 when concomitant perianal disease is present.

Table II. Summary of the large monozygotic (MZ) and dizygotic (DZ) twin studies to date.

Population	**Monozygotic twins Crohn's Disease (CD) Concordance**	**Dizygotic twins Crohn's Disease (CD) Concordance**	**References**
Danish	5/10	0/27	(Orholm, Binder et al. 2000)
British	5/25	3/46	(Thompson, Driscoll et al. 1996)
Swedish	9/18	1/26	(Tysk, Lindberg et al. 1988; Halfvarson, Bodin et al. 2003)
European	*20-50%*	*>10%*	(Halme, Paavola-Sakki et al. 2006)

Table III. Overview of the Genome-Wide Association Studies (GWAS) Performed to Date in IBD

Location of the Centers	Design	No. of SNPs genotyped	Novel Gene Identified	Other Genes Identified	References
Japan, UK	Case-control (both screening and replication)	72,738	TNFSF15	TNFSF8	(Yamazaki, McGovern et al. 2005)
Germany, UK	Case-control (screening) + Trio (replication)	19,779	ATG16L1	OCTN1/SLC22A4 NOD2/CARD15	(Hampe, Franke et al. 2007)
North America	Case-control (screening) + Trio (replication)	308,332 (Illumina, HumanHap300 BeadChip)	IL23R	NOD2/CARD15	(Duerr, Taylor et al. 2006)
	Case-control (screening) + [Trio+ Case-control (replication)]	308,332 (Illumina HumanHap300 BeadChip)	10q21.1 "gene desert" (EGR2) PHOX2B NCF4 FAM92B	ATG16L1	(Rioux, Xavier et al. 2007)
Belgium	Case-control (screening) + [Trio+ Case-control (replication)]	308,332 (Illumina HumanHap300 BeadChip)	5p13.1 "gene desert" (PGER4)	NOD2/CARD15 IL23R ATG16L1	(Libioulle, Louis et al. 2007)

Table III. Overview of the Genome-Wide Association (GWA) Studies Performed to Date in IBD (CONT’D)

Location of the Centers	Design	No. of SNPs genotyped	Novel Gene Identified	Other Genes Identified	References
UK	Case-control (screening) Case-control (replication)	500,568 (Affymetrix GeneChip 500K Mapping Array Set)	IRGM NKX2–3 PTPN2 3p21 1q "gene deserts" IL12B FLJ45139	NOD2/CARD15 IL23R ATG16L1 IBD5	(Wellcome Trust Case Control Consortium 2007) (Parkes, Barrett et al. 2007)
Mixed populations	Meta-analysis Case-control and case-parent design (replication study)	Focus on 526 SNPs from 74 distinct genomic loci	PTPN22 ITLN1 IL12B CDKAL1 CCR6 JAK2 C11orf30 LRRK2 MUC19 ORMDL3 STAT3 ICOSLG	IL23R ATG16L1 MST1 PTGER4 IRGM TNFSF15 ZNF365 NKX2-3 NOD2 PTPN2	(Barrett, Hansoul et al. 2008)

Table III. Overview of the Genome-Wide Association (GWA) Studies Performed to Date in IBD (CONT'D)

Location of the Centers	Design	No. of SNPs genotyped	Novel Gene Identified	Other Genes Identified	References
North America (pediatric IBD population)	Case-control (both screening and replication)	550,000 (Illumina Human Hap550 Genotyping BeadChip	STMN3 RTEL1 TNFRSF6B ARFRP1 ZGPAT LIME1 ZBTB46	IL23R NOD2 HLA-DRB1 PSMG1 TNFRSF6B	(Kugathasan and Cohen 2008)

Section B. Rationale and hypothesis

CD is considered a complex disease with contributions from genetic, environmental and immunological factors. Evidence for genetic risk factors is now well established with respect to the NOD2/CARD15 gene. Although other specific genomic regions have been identified as playing an important role in CD, there is currently limited information on other susceptibility genes. The study of the leukotriene-generating pathway is of particular relevance, as it represents the final step in the pathogenesis of inflammation, a key process in CD. In addition to an aberrant innate and/or adaptive immune response, susceptibility to disease may reflect the inter-individual variability in the pathogenesis of inflammation, due to differences in the expression of genes that are involved in the pathway of leukotrienes production. Thus, of particular relevance are polymorphisms in genes that play an important role in this pro-inflammatory pathway. As such, both the ALOX5 gene that regulates the pathway upstream and was shown to have an important role in the pathogenesis of inflammation in chronic disorders, and the CYP4F2 gene, which is involved in the de-activation of LTB4 and is located in a region of significant linkage to CD, may be important candidate genes in the study of CD. In spite of the overwhelming support for a potential role of these genes in the pathogenesis of CD, no study has so far investigated these associations. **In this study, we have hypothesized that DNA variants in the genes implicated in the 5-LO pathway, which plays a critical role in regulating the production of inflammatory leukotrienes, could modulate the risk for developing early-onset CD and can further influence the disease phenotype.** In the context of our project, we have particularly focused on these two genes as potential candidates for CD, the ALOX5 and CYP4F2 gene, respectively, based on the available published evidence, as well as the critical points in which they intervene in the 5-LO pathway (upstream and downstream, respectively).

Objectives

Based on this hypothesis, the primary objectives of this study were:

1. In the first stage, to examine the associations between selected DNA variants in the ALOX5 and CYP4F2 genes and early-onset of CD; and,
2. In the second stage of the study, to retest potential associations between DNA variants in the ALOX5 and CYP4F2 gene observed in the first phase of the study.

In addition the secondary objectives included:

1. To explore whether risk for CD associated with the CYP4F2 and ALOX5 genes were influenced by genomic imprinting mechanisms (parent-of-origin effect);
2. To investigate whether the CYP4F2 and ALOX5 genes were associated with specific clinical phenotypes of CD such as gender, disease location and behaviour, and
3. To determine whether interactions between the ALOX5, CYP4F2 and NOD2/CARD15 predisposed to early-onset CD.

Section C. Methods

In order to address the two main objectives of this study, we implemented a 2-phase approach. In the first phase, we explored associations between SNPs in the 2 candidate genes and pediatric CD and in the second phase we investigated whether findings of potential associations detected in Phase 1 were replicated in an independent cohort.

C.1. Study designs

Case-control-studies, where sampling is conditional on the presence or absence of disease, are widely used in epidemiology for studying associations between disease and potential risk factors. In the current context, the risk factors are represented by alleles/genotypes in the candidate genes; i.e. the ALOX5 and the CYP4F2 genes. Some of the advantages in using this type of design are that it is a relatively inexpensive approach and the assessment of the "exposure" variables (i.e. genotypes) is quite straightforward. Yet, in the context of genetic studies using a case-control approach there are a few expressed concerns that can limit their validity. More specifically, from a practical standpoint, study participants, who typically provide data on exposures and other covariates, must also provide biologic material – usually blood, for genotyping purposes. Such samples are difficult to obtain, inventory and process. Concerns about potential abuses of genetic data and the procedure in itself (i.e. venopuncture to collect blood) may make genotyping healthy subjects that are used as controls and especially children hard to justify and recruit. The resulting low control participation can invalidate a study because controls who self-select to participate may not be representative of the base-population studied. An important concern regarding the use of unrelated cases and controls in association studies is that one cannot distinguish valid association due to linkage from spurious association due to confounding effects. One of the major confounders of importance in genetic association studies is population stratification, which occurs if the population from which the cases and controls were sampled consists of latent subpopulations, each with different variant allele frequencies and risks of disease. A spurious association due to this confounding effect will occur for any variant allele that is at an elevated frequency in the subpopulation with the greatest disease prevalence.

In order to overcome the limitations of traditional case-control studies for studying genetic associations, non-traditional case-control designs have been proposed such as the case-sibling design [255] and the case-parent design [256]. In general, rather than using unrelated controls, these designs utilize familial controls. The use of family-member case-control designs is appealing because family members have a common gene pool and hence the problem of population stratification is overcome by matching. From among the family-based case-control designs, the case-parent design has been most widely used. First introduced by Falk and Rubinstein (1987) to avoid spurious association from population stratification, they recommended collecting the parents of each case subject and using the non-transmitted parental alleles as a control sample. The key assumption is that transmission of alleles from parents to offspring follows Mendelian probabilities and that, for each parental mating type, the Mendelian genotype proportions persist among offspring until the ages at which the probands are studied [257]. In doing this, the cases and controls are matched in genetic ancestry and are therefore robust to population stratification.

Based on this premise, Spielman et al. constructed a joint test of linkage and association, called the "transmission/disequilibrium test" (TDT) that attempts to identify preferential transmission of alleles from parent to affected child by use of a McNemar statistic [258]. A variant of the TDT, developed by Schaid and Sommer is a likelihood procedure for triads, called the "conditional on parental genotypes" (CPG) approach, that models the probability of an affected offspring's genotype conditional on parental genotypes as a function of the genotype relative risks (RRs) of the offspring. This CPG likelihood approach allows flexible modeling of the genotype RRs, which can be estimated using standard maximum-likelihood procedures [259].

More recently, Nagelkerke et al. advocated an approximate analysis that, although having the advantage of being easily conducted by use of standard logistic-regression software, makes the strong assumptions of Hardy-Weinberg equilibrium (HWE), random mating, and a multiplicative model of allele effect on disease [260]. A modified, power-preserving version of the Nagelkerke et al. likelihood-based approach was introduced by Epstein et al. shortly thereafter, to allow for more-flexible modeling of allele effects and less-restrictive assumptions about the distribution of parental mating types and genotypes [261].

The sampling of parental controls is often more difficult and more expensive than that of unrelated controls, since studies must identify and sample the two parents of an affected subject. Many parents may not be available for analysis, because of death, refusal to participate, or false paternity. In his approach, Epstein et al. also addressed the issue of missing information. This generalized approach can also accommodate in the analysis triad samples with missing parental data, by inferring the missing genotypes based on the observed parental genotype data. In such instance, the statistical method used is an expectation-maximization (EM) algorithm that allows the incomplete triads to contribute their information to the likelihood-ratio test without invalidation of the analysis. Simulation results also suggest that using likelihood estimates for cases missing one or both parents from the analysis can recover much of the power that would be lost if we were to exclude data from these incomplete triads from the analysis [261].

In addition to overcoming potential confounding due to "population stratification", there are other advantages of using the case-parent design *vis-à-vis* the traditional case-control design. One advantage is that it enables more efficient study of haplotype-based risks in addition to genotype-based risks. Studying haplotype-based risks is considered more powerful for the detection of potential candidate genes [262]. The latter is also possible using the traditional case-control design. However, in the absence of genotype information from the parents, it is not possible to accurately determine for homozygotes the "phase" of the transmitted allele, and hence haploytpes have to be inferred statistically. Although, many methods based on the EM algorithm have been proposed to infer haplotypes in case-control data [263], the accuracy of these estimations is less than optimal and can lead to biased inferences. A second advantage of the case-parent design is that it enables the estimation of imprinting effects (i.e. parent-of-origin effects) that cannot be estimated using the case-control design. Although the case-parent design provides numerous advantages, there are also some inherent limitations. First and foremost is the inability to determine the independent effects of "non-genetic" risk factors and subsequently its inability to control for potential confounding from such "non-genetic" variables besides ethnicity.

For the present study, we took advantage of the strengths of both types of design, thereby also exerting some control over their limitations, by implementing a 2-phase design. In the first phase (exploratory phase), we implemented the case-parent design [264-266]. Using this design we examined associations between the targeted candidate genes and susceptibility for CD for estimating genetic risks (including haplotype effects) (Objective 1) and for determining parent-of-origin effects (secondary objective), i.e. imprinting. In the second phase (replication study, Objective 2), we employed a case-control study on a separate subset of pediatric CD cases (same phenotype), to confirm potential associations observed in the exploratory phase. The cases from both Phase 1 and Phase 2 were then combined to investigate gene-gene interactions (secondary objective), whether gender modified the associations between the study genes and CD (secondary objective) and whether the study genes were associated with specific clinical phenotypes of CD, i.e. disease location and behaviour (secondary objective). Details are provided below.

C.2. Study population

The study was implemented at three pediatric gastroenterology centers across Canada Ste. Justine's Hospital (HSJ) in Montreal, the Children's Hospital of Eastern Ontario (CHEO) in Ottawa and British Columbia's Children's Hospital (BCCH) in Vancouver. These three centers are the main pediatric IBD treatment centers in the respective metropolitan areas and are expected to cover the majority of the early-onset IBD population in their respective catchment areas.

C.3. Exploratory approach (Phase 1)

The case-population for the case-parent study was represented by patients diagnosed before age 20 and that were being followed up for treatment and management at the three study hospitals mentioned above. These included both prevalent (Montreal) as well as new cases (Montreal, Ottawa, Vancouver). For non-genetic studies, selecting incident cases are ideal in order to avoid biases (such as reverse causality, survival bias, migration bias etc.) inherent when using prevalent cases. For genetic studies however, as the exposure of interest is fixed (genotype) and established prior to the outcome, including prevalent and incident cases is not a concern. CD patients that were undergoing follow-up during 2003-2007 were invited and recruited for the study – ***Appendix 1***.

The majority of the cases originated from HSJ in Montreal (n=127, 79.4%), whereas 17 (10.6%) were from CHEO (Ottawa) and 16 (10%) from BCCH (Vancouver). A total of 160 complete trios (i.e. a trio is formed by a case and both its parents) were recruited for the first phase of our study.

Overall the diagnosis of CD was based on standard clinical, endoscopic, radiologic and histopathological criteria [267, 268]. These included: (i) the presence of characteristic small bowel involvement with or without obstructive symptoms secondary to stenosis or stricture; (ii) deep linear ulcers, cobblestoning and discontinuous inflammation at endoscopy; or, (iii) patchy transmucosal chronic inflammation with granulomas on biopsy specimens. For cases whose diagnosed was unclear, confirmation was acquired after a minimum follow-up of 12 months. Clinical phenotypes of the cases were characterized at diagnosis – ***Appendix 2.***

A web-based clinical database of information from the medical charts of IBD patients diagnosed and managed at the three study centers was created. The information included location and extent of disease, and disease behavior. Information on type of medications administered, their doses and duration, and surgical interventions was also extracted and maintained. Clinical phenotypes were classified according to the Vienna classification [32] and subsequently updated by the Montreal Classification [33] - **Table I** (Page 10).

The parents of the cases were contacted and invited to participate in the study. Only parents who reported being the biological parents of the child were included. In addition, only those trios where both parents were available were included. No other major exclusion or inclusion criteria were imposed.

In order to recruit the children and their parents the following strategy was utilized. For HSJ, a comprehensive list of all patients undergoing follow-up was established in coordination with the gastroenterologists. The gastroenterology outpatient clinics of HSJ maintain a list of all such patients. In order to include potential cases missing from these lists, the gastroenterologists were asked to prepare a list of patients that they currently followed-up. After acquiring these lists, the medical charts of all the patients were screened to first confirm that the patient had been diagnosed with IBD and with CD after at least 1 year of follow-up. In addition to prevalent cases, cases newly diagnosed during the time period of the study (2003-2007) were also included. As most patients clinically suspected of having IBD undergo endoscopy, all patients who on endoscopy were suspected of having CD were temporarily listed as potential subjects. Once the initial diagnosis was established (~1 month after endoscopy), these subjects were listed for recruitment.

For both prevalent and newly-diagnosed cases, the permission of the gastroenterologist was acquired to contact the patient (as required by the Ethics Committee at HSJ). At the next scheduled visit of the patient, the parents/guardians were approached and invited for participation. For those patients (usually prevalent subjects with long-standing disease) for whom a clinic visit was not scheduled for in the near future, the contact information of these patients was sought. A letter describing the modalities of the study was sent to these families and they were informed that a research co-ordinator would subsequently contact them for soliciting participation. Approximately 10 days subsequently, the parents (usually the mother) of the patients were contacted by phone. Details of the study were explained and upon preliminary agreement to participate, a short socio-demographic questionnaire and consent forms were mailed to their residence. Alternatively, for newly-diagnosed patients who were approached in the clinics, the study details were explained in the clinic and consent acquired.

At BCCH (2004) and CHEO (2006-2008) only newly-diagnosed patients were recruited. The same strategy utilized for newly diagnosed patients at HSJ was implemented at these centers.

C.4. Replication study (Phase 2)

In order to examine whether potential associations observed in the Phase 1 study (i.e. case-parent study) withstood replication, a case-control study was carried out.

C.4.1. Cases

Cases for this study were those recruited at two pediatric gastroenterology clinics, one in Montreal (HSJ) and the other in Ottawa (CHEO). For this phase of our study, a total of 225 pediatric CD cases were included in the analyses.

The same criteria for diagnosis as was implemented in the exploratory phase were implemented in the replication study. The cases for the replication study were selected by-and-large independent of the case-parent study. The case population nonetheless included those cases who were initially recruited as part of the first phase of the study, but for whom one or both parents were not available (~23 cases) – ***Appendix 1***. It would have been ideal to randomly split the cohort of available cases and implement the exploratory and replication studies. However, our replication study was planned subsequent to the findings in the exploratory phase and hence practical limitations precluded the use of random methods. Nonetheless we believe that the selection factors involved in the exploratory and replication cohorts were similar (and not influenced by the exposure under study) and are unlikely to influence the validity of the findings in a major way.

C.4.2. Controls

One of the main challenges in carrying out a genetic study in children and when implementing the case-control design is selection of healthy population-based controls. As these controls are expected to provide DNA, the response rates could be low and low rates could invalidate the representativeness of the controls and the study-base principle. In order to overcome the latter limitations many investigators select as controls healthy adult controls. The rationale for the selection of adult controls for pediatric genetic studies is that the gene pool is not expected to change in one generation and hence the distribution of the genes under study among adults is likely to represent the distribution of the genes in that among healthy children. However, it has been shown that the gene pool in successive generations could be altered if the population structure gets altered. Immigration, inbreeding and emigration if present in successive generations than the gene pool in children would likely differ from those among individuals in the previous generation (adults).

When selecting controls for our pediatric cases, in order to ensure representativeness of the genes under study we decided to recruit controls from different sources. One source of controls was children visiting the orthopaedic departments of the study hospitals. At both study hospitals, computerized listings of all patients visiting the Acute Trauma Centers (ATC) of the orthopaedic departments are maintained. We utilized this source for selection of controls. Based on the sampling frame we selected as controls children who were diagnosed with minor traumas (viz. fractures). In addition, all controls that were concurrently diagnosed with immune mediated diseases such as asthma, juvenile diabetes, celiac disease etc. were excluded. Whenever possible we selected controls such that there was more than 1 control per case. To enhance sample size and population representativity we also invited the siblings of the orthopaedic controls to participate in the study.

In order to further ensure the representativeness of the control population and also to examine whether adult controls were appropriate for pediatric genetic studies we included as controls a population of adult controls that were being recruited for an ongoing genetic epidemiology study at the Montreal study center. The study explores the population genetics of a founder effect in the Saguenay-Lac-Saint-Jean region in North Quebec. As part of the study genetic material is collected and banked. We utilized this source in order to acquire the requisite adult controls for the study. As an additional source of controls we included umbilical cord samples from a birth cohort that is currently being assembled at HSJ – ***Appendix 1***.

For the orthopaedic controls (selected for cases recruited at Ottawa and Montreal) the same procedures were implemented. A research co-ordinator at each center based on information provided by the computerized listing sent a letter to the potential control (in most cases the mother) informing them in brief about the study and intimating them on their possibility of being contacted to solicit participation. One week after the letter was mailed to them, the potential participants were contacted over phone to solicit their participation. Details on the study objectives were provided. If verbal consent was obtained, an appointment was scheduled during their subsequent visit to the orthopaedic clinic. If a visit was not scheduled than a study kit (comprising of the consent form, brief socio-demographic questionnaire and a kit for collecting saliva) was mailed to the participants. If a clinic visit was scheduled, the participants were approached at the clinic, consent taken and information required for the study collected. For the adult controls, DNA samples that were housed in DNA bank of the respective investigators was provided to us. Similarly, umbilical cord samples were provided by an investigator (and a co-investigator in the present study) responsible for establishing the DNA bank of the birth cohort at HSJ.

A total of 330 controls (i.e. 225 orthopaedic controls; 20 controls from the birth cohort, and 85 healthy individuals) were selected for the replication phase of our study. Further details on the control population are presented in **Table XIV** (*Results* section).

C.5. Genetic information

C.5.1. Biological material

In order to extract DNA, either peripheral blood (PB) samples or saliva were acquired from all participants (i.e. cases, parents, and controls). For the cases, PB for research purposes was collected during the patients' regular follow-up visit, at the same time as blood samples were to be acquired for diagnostic and follow-up purposes. If blood samples were not being collected, than saliva was collected using ORAGENE KITS® (DNA Genotek Inc., Ottawa, Canada). A research coordinator demonstrated how the saliva was to be collected. For the parents and the controls either blood or saliva was collected. For subjects (cases and parents) who were not scheduled for a follow-up visit, the ORAGENE KITS® were mailed to them. The families were contacted by phone to verbally explain the procedure for collecting the saliva samples. Weekly reminders were made by telephone to ensure that samples were collected and mailed back to the study center (i.e. HSJ in Montreal). Participants were also advised to contact our study coordinator if they had any queries related to the provision of the study material.

All study material (i.e. questionnaires and biological samples) were coded pairwise to maintain the anonymity of the participants.

C.5.2. DNA extraction and genotyping

The DNA in saliva comes from buccal epithelial cells and white blood cells found in the oral cavity. The ORAGENE® kits are extremely user-friendly and only require collecting saliva via expectoration. In addition to its effectiveness and non-invasiveness, this method provides high quality DNA of large quantity. These kits enhance the acceptance level by subjects and are an efficient alternative to collecting DNA from blood samples [269-271]. On receipt the kits were kept stored at 4ºC pending DNA extraction, according to the manufacturer's protocol. Subsequently, DNA extraction was performed according to the extraction methods provided by the manufacturer (http://www.dnagenotek.com).

For PB samples, DNA extraction was carried out using commercially available kits, i.e PUREGENE DNA® isolation kit (Gentra Systems, Minneapolis, USA) according to their validated methods (http://www1.qiagen.com/literature). Extracted DNA was subsequently quantified by spectrophotometer (NanoDrop Technologies, Delaware, U.S.A.), transferred on 96-well plastic plates and kept frozen (-80ºC) pending genotyping.

Genotyping was done using the GenomeLab Sequenom technology which is based on a newly developed genotyping assay termed iPLEX for use with the Mass ARRAY® platform. Briefly, a primer extension assay is used to perform highly multiplexed genotyping of single nucleotide polymorphisms (SNPs) present in genomic DNA amplified by a multiplex PCR. The assay uses matrix-assisted laser desorption ionization time-of-flight mass spectrometry to accurately measure the masses of short oligonucleotide primers extended by a single dideoxynucleotide. The multiplexed genotyping assays rely on the natural molecular weight differences of DNA bases. By careful analysis of primer composition complementary to the target, or by judicious addition of one or more non-complementary 5' bases to the genotyping primers, mass spectra of interleaved genotyping products can be generated with no ambiguity in allele assignment [272]. This assay is specifically designed for high-throughput SNP genotyping and can also accommodate certain insertions/deletions.

The GenomeLab™ Sequenom Genotyping System offers multiple advances including: flexible and efficient assay design (i.e. 96% success rate), improved call rates (i.e. 85%) and accuracy (i.e. error rate is less than 0.5%) and, significantly reduced cost per genotype (i.e. below 5 cents per genotype). Genotyping was performed using the high-throughput facilities at the McGill University Genome Quebec Innovation Center in Montreal (http://www.genomequebecplatforms.com/mcgill/home/index.aspx).

C.5.3. Selection of SNPs in the CYP4F2, ALOX5 and NOD2/CARD15 genes

SNPs are the commonest variations across genes and genomes. On average, about 40-50 SNPs can be found per gene. Since many of these SNPs can be in linkage disequilibrium (LD)[***], information acquired from genotyping a few relevant SNPs can provide information on neighbouring SNPs. That means that if a causal polymorphism is not genotyped, we can still hope to detect its effects through LD with polymorphisms that are typed. Statistical approaches for identifying these “tag”-SNPs are well developed.

We used the approach proposed by Carlson et al [273]. In brief, this method is based on LD between SNPs. By re-sequencing the gene in a random sample of subjects, SNPs are identified. Subsequently using a defined threshold for LD (i.e. r^2=0.80 or more) and specifying the minimum allele frequency (MAF) of interest (i.e. 10%, for example), SNPs in LD with each other are grouped into subsets, called bins. Those SNPs with the highest LD with other SNPs in the bin are identified (or tagged) after an iterative process using a greedy algorithm. The tagged-SNPs are the ones that are genotyped. Only 1 tagged-SNP needs to be selected per bin. The set of SNPs so selected are good representatives of the un-assayed SNPs.

Besides being cost-effective, with relatively stringent LD specification, more than 80% of the haplotypes existing in the population can be resolved, enabling the study of haplotype-based associations. Public databases are now available that provide raw sequencing data for different genes among different populations that can be utilized for identifying tag-SNPS. Using the LDSELECT® program (http://pga.gs.washington.edu), 14 tag-SNPs in the 2 candidate genes were identified ***(Appendix 3***), when using genotyping data available for European Caucasians for genotyping. To identify the latter, we used the Seattle SNP public database (http://pga.gs.washington.edu) that provides information on SNPs prevalent in populations of European descent and enables

[***] Linkage disequilibrium is a term used for the non-random association of alleles at two or more loci, not necessarily on the same chromosome.

identification of tag-SNPs based on implementation of the Carlson et al algorithm [273]. The tag-SNPs were derived from DNA samples of unrelated parents (23 samples) that are part of the Centre d'Etude du Polymorphisme Humain (CEPH) repository, which is also 1 of the populations included in the HAPMAP project (www.hapmap.org). These samples are of Utah, USA, residents with ancestry from northern and western Europe. Three of these 23 samples are of subjects with French pedigrees.

As for the selection of the SNPs in the NOD2/CARD15 gene, we used publicly available information. The caspase-activation and recruitment domain gene 15 (i.e. NOD2 or CARD15 gene) has been identified as the candidate gene on the IBD1 susceptibility locus (on chromosome 16q12) and confers risk for development of CD [82]. Its 1,040 amino acids are structured into four distinct domains: two N-terminal CARDs, a central nucleotide-binding and oligomerization domain (NBD or so-called NACHT domain), and a C-terminal leucine-rich repeat (LRR) domain [274]. Three variants within the LRR were initially described to be associated with CD [70]. The three main variants are highly disease specific [70, 275, 276], are located in the distal third part of the gene, and are respresented by two missense point mutations [2104C>T (rs17860491) and 2722G>C (rs17860492)] and a frameshift mutation [3020insC (rs17860493)] resulting in nonconservative variations of the protein (respectively, R702W, G908R, and 1007fs). Altogether, these three disease susceptibility alleles represent 82% of the mutated chromosomes in CD [140]. These three disease susceptibility variants in the NOD2/CARD15 gene, identified as SNP8, SNP12 and SNP13, were genotyed in the cases (combined) and the controls. Genotyping was carried out using the SEQUENOM platform at the McGill University Genome Quebec Innovation Center in Montreal. In our sample population, individuals were categorized into NOD2+/NOD2-, depending on whether they carried any one of the three disease susceptibility variants.

C.5.4. Quality control and evaluation of genotyping errors

The genotyping process involved a stringent protocol for maintaining the quality of the results acquired. Evaluation of genotyping errors was carried out by examining Mendelian transmission patterns (case-parent design) and testing for HWE (case-parent and case-control designs). As the distribution of alleles among the alleles non-transmitted from parents to the child (pseudo-controls) is considered to be a measure of the population-allele frequency, for the case-parent study HWE was evaluated among the pseudo-controls formed by the complete trios. Thus for each trio, the allele not transmitted from the mother and father (i.e. a pair of alleles) were used as pseudo-controls. For the case-control design HWE was evaluated among the controls.

As departures from Mendelian transmission patterns could be due to false paternity or genotyping errors, distinguishing between these would be important. Douglas et al. (2002) have proposed algorithms to enable identifying whether Mendelian inconsistencies are likely the result of false paternity. In general, for every 100 families that are genotyped for ~20 SNPs, with expected genotyping error rates of ~5%, false paternity is likely to exist when for any family there are more than 6 inheritance inconsistencies (i.e. inconsistencies for 6 SNPs) [274].

We applied a modified version of these algorithms to our trio samples to detect and exclude families demonstrating false paternity. In the absence of false paternity, departures from Mendelian transmissions were considered due to genotyping errors. The SNPs that demonstrated departures from Mendelian transmissions (i.e. >5% of the families) and that deviated from HWE (i.e. $p<0.01$, accounting for the number of SNPs being genotyped) were considered candidates for replacement with other tag-SNPs.

In addition, to assess the reproducibility of the genotyping techniques, 20% of the samples were randomly re-genotyped for all the SNPs [275].

C.6. Sample size and power

C.6.1. Phase 1 - sample sizes for case-parent design

For estimating main effects using the case-parent design, the following parameters were assumed: (1) allele frequencies between 10%-50%; (2) power of 80%; (3) an alpha level of 0.05 and, (4) relative risks between 1.5 and 2.0. ***Appendix 4a*** outlines these power calculations. Based on the above parameters, it was estimated that between ~150-200 trios would be required to detect the expected risks with sufficient power assuming either a multiplicative or a dominant model. For most SNPs, power under the recessive model would not have been achievable unless very high relative risks were assumed. Our aim was thus to have sufficient power to detect risks based on either the multiplicative or dominant model while assuming modest risks for disease.

C.6.2. Phase 2 - sample sizes for case-control design

The case-control study was designed subsequent to observed findings of the case-parent study (exploratory study). To confirm the observed associations between the three SNPs identified in the exploratory study, 225 cases were available. We anticipated selecting 1-2 controls per case (average case-control ratio of 1.5). As the case-control design was carried out subsequent to the findings of the case-parent design, most of the parameters for estimating power were known. Based on these parameters (allele frequencies for the three CYP4F2 SNPs (rs1272, rs3093158, rs3093145) of, 22%, 35% and 45% respectively, odds ratios ranging from 1.5 to 4.0 (based on either the multiplicative or dominant model, 1 sided p-values of 0.05 and power=0.80), the 250 cases (x ~1.5 controls) would have more than 80% power to detect significant associations if present – ***Appendix 4b***.

As the estimation of interaction effects, parent-of origin effects and genotype-phenotype relationships were the secondary objectives and exploratory in nature, no *pre-hoc* sample size estimations were made. In order to enhance power however, for investigating gene-gene interactions, gender-gene and gene-phenotype interactions, we combined the cases from the exploratory and replication studies and compared them with the controls selected for the replication study.

Power estimations were made using the QUANTO® software (free download from http://hydra.usc.edu/gxe).

C.7. Statistical analyses

C.7.1. Disease models

For complex traits, such as CD, the mode of inheritance is unknown. In our study, we modeled disease risks assuming three potential modes of inheritance of CD: multiplicative, dominant or recessive. When a dominant model is considered, it is assumed that carriers of either 1 or two risk alleles have the same risk for disease. Alternatively, in the recessive model, it is assumed that individuals have to carry both risk alleles in order to express disease and that those carrying 1 allele have the same risks as those not carrying any variant allele (wild-type homozygotes). An alternative perspective to analyse case control data is by estimating allelic risks. This method breaks down genotypes to compare the total number of alleles in cases and controls (or pseudocontrols), regardless of the genotypes from which these alleles are constructed. This analysis provides the most powerful method of testing under a multiplicative genetic model, where risk of developing a disease increases by a factor r for each risk allele carried: risk r for heterozygotes and r^2 for high-risk homozygotes.

C.7.2. Case-parent design (Phase 1)

C.7.2.1. Independent effects of SNPs and haplotypes

Genotyping data was checked for deviation from Mendelian transmission and departures from HWE prior to analysis. For single marker analyses we used the CPG approach developed by Schaid and Sommer [259, 276]. The method does not require the assumption of HWE and similarly can model, in addition to the multiplicative model of inheritance, other possible models such as recessive, and dominant enabling the selection of a model that best fits the data.

Genotype relative risks, corresponding 95% confidence intervals and p-values were estimated for each SNP in the 2 candidate genes.

In addition to single SNP analysis, we also carried out haplotype analyses using the methods described by Purcell et al (2007) [277]. These methods are extensions of recent methods described by Schaid et al (2002) [278]. A weighted maximum likelihood model was utilized to account for the potential ambiguity in the individual's statistically-inferred haplotypes. For H number of haplotypes, two types of basic tests were carried out. A 'H' number of separate, 1 degree of freedom (df) tests of each specific haplotype compared to all others; and a single omnibus H-1 df test, jointly testing all haplotypes. Both parents were phased separately. Assuming independence between paternal and maternal chromosomes and based on basic Mendelian laws the probability of each offspring phase consistent with the observed offspring genotypes and parental phase were estimated. The contribution to the likelihood from each individual was parametrized using logistic regression and maximum likelihood was used to estimate the regression coefficients.

In addition to the above 2 tests (i.e. omnibus and haplotype-specific test including all haplotypes), we also carried out haplotype analysis including only those SNPs that showed significant association in the single SNP analysis. All haplotype analyses were carried out using the WHAP® software developed by Purcell et al. [277] - http://pngu.mgh.harvard.edu/~purcell/whap/future.shtml.

C.7.2.2. Parent-of-origin effects

Parent of origin effects were examined by applying the TDT to data stratified according to whether the variant allele was transmitted from either the father or mother. Differences in transmission between parents were assessed using the z-test and p-values were estimated. Analysis was carried out using the TDT parent-of-origin feature in PLINK® - http://pngu.mgh.harvard.edu/purcell/plink/.

C.7.3. Case-control study (Phase 2)

The major objective of the case-control study was to confirm the findings of association in the exploratory study. For this study all SNPs that were significantly associated at the allelic or genotypic level (individual SNP analysis) with CD in the exploratory study (case-parent design) at $p \leq 0.10$ were subjected to replication.

Initial univariate analysis was carried out to study the distribution of the genotypes and alleles among cases and controls. Allelic associations and genotype associations were investigated. For allelic associations odds ratios and respective 95% CI were estimated. Genotype associations were evaluated by applying chi-square tests to the 3 x 2 contingency table. Further analysis was carried out using unconditional logistic regression. Three different models of inheritance were examined: (i) model where the risks for heterozygotes and variant homozygotes were considered distinct, (ii) a recessive model, and (iii) a dominant model. The different models were examined using the Akaike's Information Criteria (AIC), which is a proposed measure of the goodness of fit of an estimated statistical model [279]. Models with the lower AIC were considered to best fit the data. Each SNP was examined in separate models. Odds ratios and respective 95% CI were estimated.

Haplotype analyses were carried out using the methods described by Purcell et al (2007) [280]. For 'H' number of haplotypes, two types of basic tests were carried out. A 'H' number of separate, 1 degree of freedom (df) tests of each specific haplotype compared to all others; and a single omnibus H-1 df test, jointly testing all haplotypes. All haplotype analyses were carried out using the WHAP® software developed by Purcell et al. [280] - http://pngu.mgh.harvard.edu/~purcell/whap/future.shtml.

C.7.4. Examining gene-gene interactions

To assess gene-gene interactions, we combined the case population from the exploratory study as well as that from the replication study.

Assessing gene-gene interactions using moderate sample sizes may limit the power to detect important interactions. In addition considering the multiple numbers of tests that are carried out, required adjustment for multiple testing can further limit power. These limitations are more so in parametric methods such as logistic regression wherein at higher dimensions the data becomes sparse. High-dimensional contingency tables lead to many cells with no or small number of data points leading to large standard errors and thus high type-1 errors. Similarly in parametric methods genes that do not have independent effects are usually not selected (for e.g. in step-wise logistic regression) and hence their contribution to gene-gene interactions is likely to be missed. In order to overcome these limitations, data reduction strategies such as multi-factor dimensionality reduction (MDR) have been proposed [280, 281] to detect gene-gene interactions in the presence or absence of main effects in case-control studies in human genetics [161, 282]. MDR is a non-parametric genetic model-free approach wherein multi-locus genotypes are grouped into "high-risk" and "low-risk" groups thus reducing the N-dimensional space into one dimension. This is achieved through a stepwise process. In step 1, an exhaustive list of n combinations of genetic loci to evaluate from the list of all variables is created. Next, each of the n combinations is arranged in contingency tables in k-dimensional space with all possible combinations as individual cells in the table. Then, the number of cases and controls for each locus combination is counted. In step three, the ratio of cases to controls within each cell is calculated and each genotype combination is labeled as "high-risk" or "low-risk" of the phenotype of interest based on comparison of the ratio to a threshold. The threshold used is dependent on the ratio of cases and controls within the data set. If the ratio within a multifactor combination is above that seen in the data, it is labeled as “high-risk” and if it is below, it is labeled “low-risk”. This step compresses multidimensional genotype data into one dimension with two classes. The new multi-locus variable is evaluated for its effectiveness in discriminating between the presence (case) and absence (control) of disease using cross-validation and permutation testing. Because the MDR algorithm is using permutation distribution for significance testing, correction for multiple testing is unnecessary. Using simulated data, it has been shown that MDR methods have much greater power than traditional methods such as logistic regression and classification and regression trees (CART) in detecting gene-gene interactions and particularly detecting genes that do not have main effects [283].

In our exploratory study, we examined gene-gene interactions between three SNPs in the CYP4F2 gene (that were suggestive of associations in the exploratory study), four SNPs in the ALOX5 gene, and the three disease susceptibility variants (SNP8, SNP12 and SNP13) in the NOD2/CARD15 gene. For the NOD2/CARD15 gene, individuals were categorized into NOD2+/NOD2-, depending on whether they carried any one of the three disease susceptibility variants, as previously explained. The entire data were randomly classified into two subsets (9/10 and 1/10) with the larger subset (9/10) being the training sample and the smaller subset (1/10) providing the testing sample. The MDR procedure was run in both the training and the testing samples and the accuracies of the model were examined. The procedure was run 10 times (i.e. 10 cross validations). The results for the 10 cross-validations were averaged to acquire the testing accuracies or prediction errors (which is 1-testing accuracy). Based on the procedure the best model (based on the testing accuracy), separately for each order of interaction (i.e. best models for a 2, 3, 4 order interaction etc.) is identified. In order to limit over-fitting of the data that can occur with high-order interactions (viz. 4) we restricted our analysis to 2 and 3 order interactions.

Significance of testing accuracies of the models selected using the MDR procedure was estimated based on permutation methods. In order to carry out the procedure, the case/control labels of the study data were randomly permutated to create 1,000 data sets. The same MDR models were re-run on the 1,000 permuted data sets and the testing accuracies of the best models were sorted in ascending order. The p-value for an observed model was the proportion of times the best testing accuracy of the permuted data was higher than the observed testing accuracy. Thus if the testing accuracy of the observed model was 0.59 than a p-value of 0.05 would indicate that 5% of the testing accuracies from the 1,000 permuted data sets were higher than the observed testing accuracy. As the permutation process is carried out using the same procedures and same models utilized in the MDR procedure, the acquired p-values are automatically adjusted for multiple testing [281].

All analyses were carried out using the MDR version 2.0 (open-source software) and related utilities (MDR-permutation testing) (http://sourceforge.net/projects/mdr/).

C.7.5. Examining gender-gene and genotype-phenotype associations

Gender-gene interaction analysis was restricted to the CYP4F2 gene. Only 1 SNP in the CYP4F2 gene withstood replication and hence gender-gene interactions with this SNP were tested (viz. rs3093158). Cases from both the exploratory and replication studies were combined and compared to the unrelated controls, i.e. 385 cases versus 330 unrelated controls. Interaction analysis was carried out by fitting a logistic regression model including the CYP4F2 genotype (coded according to a dominant model), gender and an interaction term involving the genotype and gender to the binary outcome (case-control status). Significant interactions were deemed to be present when the interaction term was statistically significant at the 5% level.

The two main phenotypes of CD, disease location and disease behaviour at diagnosis were the phenotypes evaluated in the study. Both a case-based analysis and a case-control analysis wherein the allele frequencies within a particular phenotype was compared with that among controls was carried out. The case-only analysis was based on the combined set of cases (160 plus 225), whereas the case-control analysis was based on the combined set of cases (i.e. 385) versus the unrelated controls (i.e. 330). Based on the Montreal classification disease location is categorized as L1±L4 (i.e. ileal with or without upper tract involvement); L2±L4 (i.e. colonic with or without upper tract involvement) and L3±L4 (i.e. ileo-colonic with or without upper tract involvement) – **Table I** (Page 10). For the purpose of the analysis, we combined patients who had ileal disease (with or without upper tract or colonic involvement) i.e. L1±L4 and L3±L4, as pure colonic disease is expected to differ genetically from ileal disease. For disease behaviour, the phenotype according the Montreal Classification is: B1±B4 (i.e. inflammatory with or without perianal involvement), B2±B4 (stricturing with or without perianal involvement) and B3±B4 (penetrating with or without perianal involvement) – **Table I** (Page 10). As B2 and B3 represent non-inflammatory disease we combined these two categories for the analysis.

In the case-based analysis, genotype associations were examined by using a chi-square test (2 df) applied to the 2 x 3 contingency tables. Allelic associations were examined by applying a chi-square test to the frequency distribution of the variant allele in the case-chromosomes versus that among the control chromosomes (1 df). For the latter analysis, odds ratios and respective 95% CI were estimated. For the case control analysis, the allele frequencies within each phenotype were compared with that among controls using chi-square tests. Odds ratios and 95% CI were estimated.

C.7.6. Linkage disequilibrium (LD) analysis

We also evaluated the pair-wise LD between SNPs in the two candidate genes. Among the measures that have been proposed to evaluate LD, the two most important are D' and r^2 [284]. The first measure, D′ is sensitive to even a few recombinations between the loci since the most recent mutation at one of them.

However, the literature emphasizes the exponential decay over time of D′ between linked loci under simple population genetic models, but stochastic effects may suggest that this theoretical relationship is of limited usefulness. A disadvantage of D′ is that it can be large (indicating high LD) even when one allele is very rare, which is usually of little practical interest. The second measure of LD, r^2 reflects the statistical power to detect LD. Therefore, a low r^2 corresponds to a large sample size that may be required to detect the LD between the markers. If, for example, disease risk is multiplicative across alleles, and HWE holds, r^2 between a marker and a causal SNP gives the sample size that would have been required to detect the disease association by directly typing the causal SNP, relative to the sample size required to achieve the same power when typing the marker.

For our study, both measures of LD, i.e. D' and r^2 were estimated. LD was determined both for the case-parent study [among the alleles not-transmitted to the index case (pseudo-controls)] and the case-control study (in controls).

All analyses were performed by STATA version 8.0 (http://www.stata.com/).

Section D. Results

All figures and tables in the Results section are presented at the end of this section (Pages 82 -106).

D.1. Case-parent design

D.1.1. Demographic and clinical data

As mentioned above, the cohort for the study was established based on CD cases (prevalent and newly-diagnosed) from HSJ (Montreal) during 2003-2007, cases newly diagnosed during 2004 at BCCH and those newly diagnosed during 2006-2007 at CHEO. A total of 160 complete trios (i.e. a trio is formed by a case and both its parents) were recruited.

The majority of the cases (n=127, 79.4%) were enrolled at HSJ (Montreal), whereas 17 (10.6%) were from CHEO and 16 (10%) from BCCH. From among the 160 patients with CD, 85 were boys (53.1 %) and 75 were girls (46.9%). Mean age at diagnosis was 11.8 years (SD = 3.3), ranging from 3 to 20.3 years (1 patient who was diagnosed just subsequent to completing 20 yrs was retained). All the patients were Caucasian. Disease was phenotypically classified on the basis of its location [terminal ileum (L1), colon (L2), ileocolon (L3), upper gastrointestinal tract (L4)] and behaviour [nonstricturing nonpenetrating (B1), stricturing (B2), penetrating (B3)]. The Montreal classification of CD clinical phenotypes was implemented. The demographic and clinical characteristics of the CD patients are shown in **Table IV**.

At diagnosis, the majority of the patients had ileo-colonic disease (i.e. 57.5%) and had inflammatory behavior (i.e. 88.8%).

D.1.2. Genetic analyses

In order to carry out a comprehensive assessment of the role of the ALOX5 and CYP4F2 genes we utilized the haplotype-based gene-tagging approach previously described. Using an r^2 cut-off of 0.80 for linkage disequilibrium (LD) between any 2 SNPs, and setting the minimum allele frequency (MAF) to 10% we identified the set of SNPs that predicted >80% of the haplotypes formed by the gene. The location of the fourteen SNPs, i.e. ten in the CYP4F2 gene and four in the ALOX5 gene, is illustrated in **Figures 6 and 7,** respectively. **Table V-A and B** shows the primers utilized for genotyping the 14 tag-SNPs in both settings, i.e. case-parent and case-control design. The chromosomal and gene locations of the SNPs are shown in **Table VI**.

In order to meet one of the main objectives of the study, i.e. to screen for associations between DNA variants in the ALOX5 and CYP4F2 genes and early-onset of CD, we implemented a case-parent design. Using this setting, fourteen tag-SNPs, 10 in the CYP4F2 gene and 4 in the ALOX5 gene, were genotyped among 182 trios. Overall the genotyping success rate was ~99%. Genotyping was successful for all three members of 179 trios. Based on published recommendations, approximately 20% of the DNA samples were genotyped in duplicate for all the 14 SNPs [278]. Samples that were not reproducible for any SNP were excluded from the analysis. This resulted in the exclusion of 3 additional trios, leaving 176 complete trios available for further analysis.

When patterns of Mendelian inheritance were examined, ten trios showed deviation from Mendelian patterns across more than 4 SNPs, suggesting that the deviations were due to mis-paternity. These trios were excluded. A further 6 trios showed deviations for 1-4 SNPs. These deviations were regarded as potentially due to genotyping errors and these trios were also excluded from further analysis leaving 160 trios for statistical analysis.

For the complete trios conforming to Mendelian patterns of inheritance, tests for deviation from HWE were carried out among the alleles not-transmitted to the case (pseudo-controls). All 14 SNPs were in HWE - **Table VII**.

Further statistical analysis of the data was thus based on 160 complete trios **(Table VIII).**

We also established the presence of linkage disequilibrium between SNPs in the two study genes among the alleles not-transmitted to the index case (pseudo-controls). **Table IX** (**a** and **b**) show the pair-wise LD structure for SNPs belonging to the two candidate genes (D' and r^2 measures).

The maximum LD for gene CYP4F2 was observed between SNPs rs1272 and rs3093198 (r^2=0.060). A r^2 of 0.56 was observed between SNPs rs1272 and SNP rs2016503. In general the LD between most of the pair-wise LD was low. This was expected as the selection of tag-SNPs using the Carlson algorithm was based on an r^2 of ≥ 0.80 between SNPs, hence none of pair-wise r^2 were ≥=0.80.

D.1.2.1. Single marker analyses

Single marker analysis was carried out using the CPG method implemented in the SCOUT® software (http://www.genetics.emory.edu/labs/epstein/software/scout/). Three models of inheritance, i.e. dominant, multiplicative and recessive, were examined.

The results of the single marker analyses for CYP4F2 gene in complete trios are presented in **Table X**.

From, among the 10 SNPs for the CYP4F2 gene, significant associations ($p<0.05$) assuming either the multiplicative, recessive or dominant models of inheritance were observed for SNPs rs3093158 (OR=4.02, p_{dom}= 0.00003), rs1272 (OR=3.10, p_{dom} =0.04), and with rs3093145 (OR=1.76, p_{dom} =0.017).

The results of the single marker analyses for ALOX5 gene are shown in **Table XI**.

None of the four SNPs in the ALOX5 gene were significantly associated with CD assuming any model of inheritance.

D.1.2.2. Haplotype analyses

By using regression-based methods, we further examined whether haplotypes formed by the tag-SNPs in the CYP4F2 and ALOX5 were associated with CD.

For CYP4F2 haplotype analyses, we included first all ten biallelic markers in our analysis. The ten SNPs formed eight haplotypes (i.e. H1-H8) in the population.

Table XII shows the haplotype descriptions, their frequency, and the p-value of a regression-based analysis of the association between each haplotype and CD.

Two haplotypes – H6 (6%) and H7 (5%), respectively - showed significant associations with CD (p=0.05 and 0.001, respectively). An omnibus test examining overall haplotype associations was also significant (p=0.035).

Subsequently, we performed haplotype analyses using the three SNPs, which showed significant associations when using the multiplicative, recessive or dominant model in the single marker analysis (i.e. rs3093158, rs1272, and rs3093145). These three

SNPs formed four haplotypes (i.e. H11-H14). When tested individually, significant associations were observed with two haplotypes (i.e. H11 and H14, respectively) that account for approximately 60% of potential haplotypes (p=0.045 and 0.004, respectively). The corresponding omnibus test for haplotype association showed strong evidence for association with CD (p=0.007).

D.1.2.3. Parent-of-origin analysis

Recent findings suggest that parent-of-origin effects may be at play in CD. These observations provided the impetus to explore such effects in the present study. **Table XIII** contains the results of the parent-of-origin analysis in the 10 SNPs of the CYP4F2 gene. In the parent-of-origin analysis, the TDT was applied separately to trios where the risk allele was transmitted to the child from the father and to trios where the transmission occurred from the mother. There seems to be an indication of association with disease, when the risk allele of two SNPs was transmitted from the fathers (rs1272: p=0.07; rs3093158: p=0.02). However, the z-test comparing the differences in transmission by parent was not significant for any of the 10 SNPs suggesting that there were no parent-of-origin effects. Interestingly, the rs3093158 SNP shows similar associations irrespective if the risk allele is transmitted from the father (p=0.02) or the mother (p=0.07), suggesting that this SNP may be associated with CD, irrespective of the parent-of-origin effects.

D.2. Case-control design

In order to address the second objective of the study, i.e. to confirm the findings from the case-parent analyses, as well as to explore gender-gene and gene-gene interactions, we implemented a case-control design, whereby we analyzed the three tag-SNPs in the CYP4F2 gene that were found to be significantly associated (i.e. p<0.05) with CD in the screening phase (i.e. case-parent design) in a separate cohort of CD cases (N=225) and controls (N=330).

D.2.1. Demographic and clinical data

The cases and controls for the case-control design originated only from HSJ and CHEO. For this phase of our study, a total of 225 CD cases and 330 controls were included in the analyses.

The majority of the cases (n=173, 76.9%) were recruited from HSJ (Montreal), whereas 52 (23.1%) were recruited from CHEO (Ottawa). From among the 225 patients with CD, 138 were boys (61.3 %) and 87 were girls (38.7%). Mean age at diagnosis was 12.4 years (SD = 3.4), ranging from 3 to 21 years. All the patients were Caucasian.

At diagnosis, the majority of the patients had ileo-colonic disease, L3 (i.e. 56%) and had inflammatory behavior (i.e. 84.8%). From HSJ, 282 (84.4%) controls were recruited, whereas 48 controls (14.6%) were recruited from CHEO. Among the controls, there was a preponderance of males (n=199, 60.3%). The mean age of the controls was 10.6 years (SD=7.2).

The demographic and clinical characteristics of the CD patients versus the controls are shown in **Table XIV**.

D.2.2. Genetic analyses (replication study)

Cases and controls were genotyped using the same methods as in the case-parent design. A full description of the primers used is presented in **Table VI**.

The results of the case-control analysis are presented in **Table XV (A, B, and C)**.

From among the three SNPs that suggested associations in the screening study, associations were replicated with SNP rs3093158 in the Phase 2 study. Significant genotype associations (uncorrected p=0.03, one-sided) were noted. After applying the standard Bonferroni correction for multiple testing, this genotype association however did not withstand correction (corrected p=0.09). When a dominant model was examined, individuals carrying 1 or more of the common variant in this SNP were at elevated risks for CD (OR=1.94, p_{dom}=0.023) even after correction for multiple testing– **Table XV-B**.

There was no evidence for association between SNPs rs1272 and rs3093145 in the case-control setting.

The results of the haplotype analyses in the case-control study are presented in **Table XVI**. The three SNPs in the CYP4F2 gene formed 4 haplotypes (H1-4). The frequencies of these haplotypes in the cases versus the controls are presented in the same table. One of the haplotypes, i.e. H4, was found to be significantly associated with CD (p=0.02).

D.3. Combined analyses

D.3.1. Combining results from the case-parent and case-control analyses

This entire experiment, i.e. a multistage approach, must be viewed as the result of both stages for valid statistical inference. This entire experiment, i.e. a multistage approach, must be viewed as the result of both stages for valid statistical inference. According to Skol (2006), a joint analysis is more powerful than treating the second stage as a stand-alone replication sample [288]. In order to gauze the significance of these results, we combined the findings of the case-parent and case-control designs by using a likelihood-based approach described by Epstein et al [288]. In short, Epstain et al developed combined tests of association that are based on the likelihood of Nagelkerke et al. that looks for overlap in CIs for the separate relative risks parameters [260]. In contrast to the latter, the method proposed by Epstein et al applies selection procedures, such as the Akaike information criteria (AIC), Bayesian information criteria, or backwards selection to formalize hypothesis testing and model selection. Backward selection can be used by starting with the richest possible model and removing parameters after hypothesis testing. Under the assumption that at least one hypothesis is rejected, the appropriate parameters can be constrained and the model can be refit. This procedure is repeated until no additional hypotheses about parameters can be rejected. Using simulated as well as real data, the authors demonstrated that this single combined association approach has improved power over statistical methods that analyze triads and unrelated subjects separately. Direct comparison of the genotype distribution among triad parents with that of unrelated controls is recommended since allele-frequency differences between the two samples may induce bias and therefore the information from these two sources cannot be safely combined.

More specific, the SCOUT software conducts a joint association analysis of triads and unrelated controls using a version of the CPG approach that is augmented to allow for unrelated control information. First, the program will run an unconstrained analysis that estimates the RR parameter from the triad information (RR_1) and the control information (RR_0), as separate parameters. Second, the program will run an analysis that constraints $RR_1=RR_0$ but estimates the RR parameter from the data. Finally, the program will run a null association analysis that constraints $RR_1=RR_0=1$. Using results from these three analyses, SCOUT then constructs and evaluates likelihood-ratio statistics for testing two hypotheses of interest. The first hypothesis is H_0: $RR_1=RR_0$, which is a test for assessing whether the controls can be safely combined with the triads for association analysis. Rejection of H_0 suggests that the two samples cannot be combined, so inference should be based on the CPG analysis of the triads only. The second hypothesis tested is H_0: $RR_1=RR_0=1$, which is a test of association between SNP and disease.

Performing this analysis in our data set, based on the multiplicative model, the combined p-value was 5.01×10^{-3} and under the dominant model, 1.51×10^{-4}. After applying a standard Bonferroni correction for multiple testing (for 3 SNPs), the adjusted p-values were 1.5×10^{-2} and 4.53×10^{-4} respectively.

D.3.2. Gender-gene interaction

A significant difference in gender distribution between early-onset and adult CD has been noted. Although endocrinologic differences are considered to underlie the sexual asymmetry of distribution, to date, no specific molecular mechanisms of such hormonal effects have been elucidated. In the context of our study, investigating the interaction between gender and risk alleles might provide additional insight into the pathogenesis of CD.

Based on the logistic regression analysis applied to the data from the combined cases from phase 1+2 (N=385 cases) versus the unrelated controls (M=330 controls), **Table XVII** presents the result of the interaction between the rs3093158 in the CYP4F2 gene with gender. No significant interaction was observed.

D.3.3. Gene-gene interaction

Both CYP4F2 and ALOX5 genes play a critical role in the eicosanoids pathway. ALOX5 is the enzyme responsible for initiating the process whereas CYP4F2 is involved at the termination of the pathway that leads to production of inflammatory mediators such as leukotrienes. On the other hand NOD2/CARD15 is known to be the key gene involved in innate immunity. We thus postulated that interactions between these three genes could modulate the inflammatory process and a combination of variants in these genes may impact disease. The location of the tag-SNPs that were analyzed and their frequencies in the unrelated controls is presented in **Table XVIII**.

To investigate the joint effects between the three genes, we applied MDR methods to the combined cases-unrelated control data (**Table XIX**).

In the MDR analysis, when two-factor models were fit involving the 8 SNPs (3 CYP4F2, 4 ALOX5 and NOD2), a model including SNP rs3093158 (CYP4F2) and NOD2 had the highest testing accuracy (0.5806) and the highest cross-validation success (the same model was identified as the best model in 8 of 10 cross-validations). This model was statistically significant on permutation testing (P-value between 0.004-0.005). Similarly when three-factor models were examined the model including the CYP4F2 SNP rs3093158, the ALOX5 SNP rs2115819 and the NOD2 gene was identified as the best model. This model was statistically significant (P-value between 0.006-0.007). These analyses indicate that although the ALOX5 gene did not appear to have any independent effects on the risk for CD, in combination with CYP4F2 and NOD2 it contributes significantly to risk for CD.

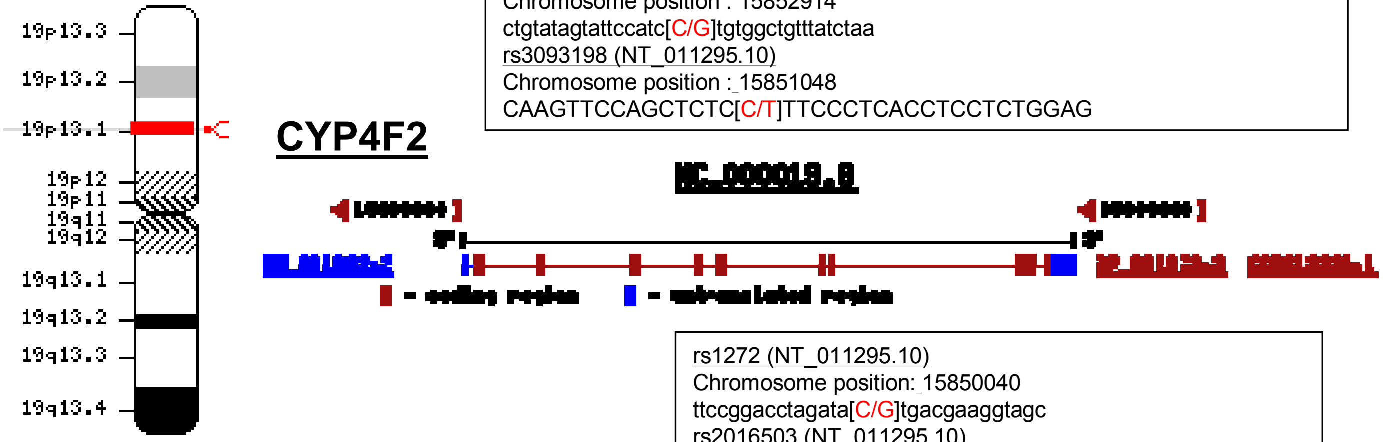
D.3.4. Case-only analyses for genotype-phenotype association

Location and disease behaviour have been shown to be two important determinants of clinical management and outcome. Patients with ileocolonic location and those with perianal involvement, for example, often suffer a more complicated course than those with disease localization in the terminal ileum or colon only. While certain locations may predict clinical outcomes, it is yet unclear why certain individuals would have predilection for disease in a certain location. In addition, a consistent difference in phenotype in children versus adults in disease distribution has been noted. Based on these observations, a secondary objective of this study was to study associations between the validated SNPs and disease phenotypes. **Table XX** contains the case-based analysis of association between CYP4F2 SNP rs3093158 with CD disease location and behaviour in Canadian children (viz. combined case population, N=385 CD cases). No significant impact of the studied SNP on disease location and/or behaviour was observed.

D.3.5. Case-control analysis for genotype-phenotype association

After combining all cases (i.e. 160 cases from the screening study + 225 cases from the replication phase), we performed a case-control analysis wherein the CYP4F2 SNP rs3093158 allele frequencies in the CD cases was compared with that among the unrelated controls (M=330). Consistent with the previous analysis, no significant impact of the studied SNP on disease location and/or behaviour was observed. The results of this analysis are presented in **Table XXI**.

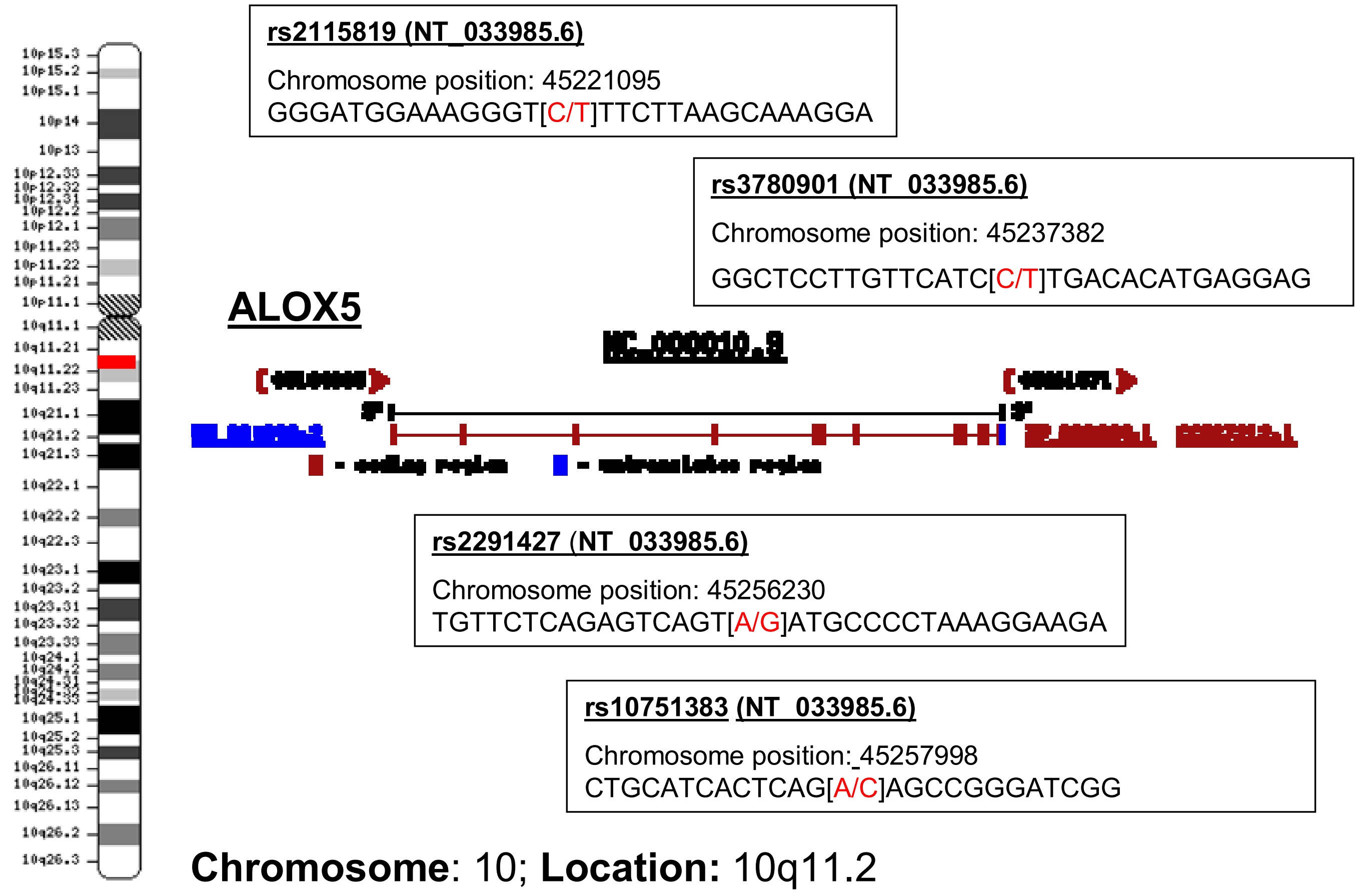
rs3093144 (NT_011295.10)
Chromosome position: 15863297
attgtggcaaaggttttgc[A/G]aagtatctaggatttt
rs3093145 (NT_011295.10)
Chromosome position: 15862852
ggactacctcccC[A/C]AAAACAGGGATATTTTTG
rs3093158 (NT_011295.10)
Chromosome position:
15861166ATGTCAGATGAAAG[A/G]ATTTGAACTTGATTAA
rs3093193 (NT_011295.10)
Chromosome position : 15852914
ctgtatagtattccatc[C/G]tgtggctgtttatctaa
rs3093198 (NT_011295.10)
Chromosome position : 15851048
CAAGTTCCAGCTCTC[C/T]TTCCCTCACCTCCTCTGGAG



rs1272 (NT_011295.10)
Chromosome position: 15850040
ttccggacctagata[C/G]tgacgaaggtagc
rs2016503 (NT_011295.10)
Chromosome position: 15866185
CCAGGAGAAGTAAA[C/T]CATGGGCCATTTCTG
rs2108622 (NT_011295.10)
Chromosome position: 15851431
AGGGTCCGGCCACA[C/T]AGCTGGGTTGTGAT
rs2074900 (NT_011295.10)
Chromosome position: 15857820
GTGCTTTGCAAG[A/G]TGGTACAGGA
rs2074902 (NT_011295.10)
Chromosome position: 15869099
GAATGAGTAAGA[C/T]GCCTCCTCTGCT

Chromosome: 19
Location: 19pter-p13.11

Figure 6 – Physical map CYP4F2 on Chromosome 19 and positioning of the SNPs



rs2115819 (NT_033985.6)

Chromosome position: 45221095
GGGATGGAAAGGGT[C/T]TTCTTAAGCAAAGGA

rs3780901 (NT_033985.6)

Chromosome position: 45237382

GGCTCCTTGTTCATC[C/T]TGACACATGAGGAG

ALOX5

rs2291427 (NT_033985.6)

Chromosome position: 45256230
TGTTCTCAGAGTCAGT[A/G]ATGCCCCTAAAGGAAGA

rs10751383 (NT_033985.6)

Chromosome position:_45257998
CTGCATCACTCAG[A/C]AGCCGGGATCGG

Chromosome: 10; **Location:** 10q11.2

Figure 7 – Physical map ALOX5 on Chromosome 10 and positioning of the SNPs

Table IV. Demographic and clinical characteristics of the CD Patients in the case-parent design (N=160)

NO. OF CASES **N=160**	
Age at diagnosis	
Mean (SD)	11.8 (3.3)
Gender (%)	
Females	75 (46.9)
Males	85 (53.1)
Study site (%)	
Montreal	**127 (79.4)**
Ottawa	**17 (10.6)**
Vancouver	**16 (10.0)**
Clinical characteristics (at diagnosis)	**At diagnosis**
Disease location*(%)	
Ileum with or without upper digestive tract, L1+/-L4	22 (13.7)
Colon with or without upper digestive tract, L2+/-L4	46 (28.7)
Ileum and colon with or without upper digestive tract, L3 +/-L4	92 (57.5)
Disease behaviour* n (%)	
Inflammatory B1 +/- p‡	142 (88.8)
Stricturing B2 +/- p	13 (8.1)
Perforating B3 +/- p	5 (3.1)

*according to Montreal classification

‡ perianal modifier

Table V-a – List of primers used to detect the 10 tag-SNPs in the CYP4F2 gene

		Position	Reference SNP	Sense	Abbreviation	Flanking Sequence
CYP4F2 LOCATION 19pter-p13.11	1.	15850040	rs1272	F	rs1272_CG_F	ACGTTGGATGCACATACCACGAAATTCACC
				R	rs1272_CG_R	ACGTTGGATGGGGATGGTGAAAATGTTCCG
	2.	15851048	rs3093198	F	rs3093198_CT_F	ACGTTGGATGCAACCCAACCGTACTCTATG
				R	rs3093198_CT_R	ACGTTGGATGGACATTGTAGATGGTCCAAG
	3.	15851431	rs2108622	F	rs2108622_CT_F	ACGTTGGATGCATCAGTGTTTTCGGAACCC
				R	rs2108622_CT_R	ACGTTGGATGCTCTAGGAGCCTTGGAATGG
	4.	15852914	rs3093193	F	rs3093193_GC_F	ACGTTGGATGGTGATGAGACTAGTGATCCC
				R	rs3093193_GC_R	ACGTTGGATGGCCACATACACATTGATGGG
	5.	15857820	rs2074900	F	rs2074900_GA_F	ACGTTGGATGATCTCTTTAGGCTCACGGTC
				R	rs2074900_GA_R	ACGTTGGATGAGTGGTCTCTCCTGGGTCCT
	6.	15861166	rs3093158	F	rs3093158_AG_F	ACGTTGGATGTCCCTTCCTCAATCACCTTC
				R	rs3093158_AG_R	ACGTTGGATGGTAGAAGGGAGCTTCATGTG
	7.	15862852	rs3093145	F	rs3093145_CA_F	ACGTTGGATGCACATGGCATTGTTTCTGGC
				R	rs3093145_CA_R	ACGTTGGATGAGGACTCAACGAAGGACTAC
	8.	15863297	rs3093144	F	rs3093144_AG_F	ACGTTGGATGAGGAGTCTCTCGTCCTTCTG
				R	rs3093144_AG_R	ACGTTGGATGGGGAAGAATTGTGGCAAAGG
	9.	15866185	rs2016503	F	rs2016503_CT_F	ACGTTGGATGCCACCTTTCCCCTAGAGTTC
				R	rs2016503_CT_R	ACGTTGGATGATCAGAGACACAGGGATTGG
	10.	15869099	rs2074902	F	rs2074902_CT_F	ACGTTGGATGTACGAGGCTTAGGGAGTGG
				R	rs2074902_CT_R	ACGTTGGATGGTAGGCACCTCACAGAAATG

Table V-b – List of primers used to detect the 4 tag-SNPs in ALOX5 gene

		Position	Reference SNP	Sense	Abbreviation	Flanking Sequence
ALOX5 LOCATION 10q11.2	1.	45221095	rs2115819	F	rs2115819_CT_F	ACGTTGGATGTGTTGCTTTTGCCACAGGAG
				R	rs2115819_CT_R	ACGTTGGATGTTTGTGTAACACTGGGATGG
	2.	45237382	rs3780901	F	rs3780901_CT_F	ACGTTGGATGAGAGGCCTTTGTGAGTACTG
				R	rs3780901_CT_R	ACGTTGGATGGGCAACTTCCAGCAGTACAC
	3.	45256230	rs2291427	F	rs2291427_AG_F	ACGTTGGATGGCTCCTAGCTGTTTTCTTCC
				R	rs2291427_AG_R	ACGTTGGATGGATCACTGACATCCCACAGG
	4.	45257998	rs10751383	F	rs10751383_CA_F	ACGTTGGATGTGAGCTTGTGAGTGTGTCTG
				R	rs10751383_CA_R	ACGTTGGATGTGTTGTGTGCTCAGGTGAGG

Table VI - Location of the tag-SNPs in CYP4F2 and ALOX5 genes

Gene/SNP		**Chromosomal location**	**Gene location**
CYP4F2 Chromosome 19pter-p13.11			
1.	rs1272	15850040	3'-utr
2.	rs3093198	15851048	intron 12
3.	rs2108622	15851431	exon 11-coding-ns
4.	rs3093193	15852914	intron 9
5.	rs2074900	15857820	exon 9-coding-ns
6.	rs3093158	15861166	intron 7
7.	rs3093145	15862852	intron 5
8.	rs3093144	15863297	intron 5
9.	rs2016503	15866185	intron 3
10.	rs2074902	15869099	intron 2
ALOX5 – Chromosome 10q11.2			
1.	rs2115819	45221095	intron 2
2.	rs3780901	45237382	intron 4
3.	rs2291427	45256230	intron 8
4.	rs10751383	45257998	intron 9

Table VII - Frequency distribution of the tag-SNPs in CYP4F2 and ALOX5 in the informative cases and the pseudo-control(s) (formed by the non-transmitted alleles, one pseudo-control versus all three pseudocontrols) belonging to the complete trios (N=160).

Gene	**SNP**	**Minor Allele Frequency (MAF) (%)**		**HWE-p-value (1/3)***
CY4F2 Chromosome Location 19pter-p13.11 Position		Cases (N=160)	Pseudo-control(s) (1/3)	
15850040	rs1272	29.4	23.9/22.8	0.28/0.90
15851048	rs3093198	35.6	28.3/27.9	0.28/0.75
15851431	rs2108622	43.5	28.4/29.1	0.46/0.72
15852914	rs3093193	42.0	36.9/36.7	0.18/0.71
15857820	rs2074900	35.9	29.5/30.0	0.14/0.26
15861166	rs3093158	32.0	38.0/35.5	0.98/0.24
15862852	rs3093145	52.2	44.2/46.2	0.24/0.60
15863297	rs3093144	34.9	14.7/17.0	0.91/0.92
15866185	rs2016503	26.6	17.1/16.8	0.87/0.80
15869099	rs2074902	32.3	17.8/17.8	0.18/0.27
ALOX5 Chromosome Location 10q11.2	**SNP**	Cases	Pseudo-control(s) (1/3)	**HWE-p-value (1/3)***
45221095	rs2115819	50.4	49.6/49.3	0.35/0.29
45237382	rs3780901	38.5	32.1/31.8	0.92/0.48
45256230	rs2291427	28.0	28.8/28.7	0.53/0.93
45257998	rs10751383	39.0	41.0/42.0	0.82/0.38

* Examined among the pseudo-controls (1 or 3)

Table VIII - Distribution of case-parent families in the study.

Case-parent families	**Trios**	**Total number of participants**
Recruited	182	546
After successful genotyping	179	537
After testing for reproducibility	176	528
After testing for Mendelian inheritance	160	480

Table IX-a – Linkage disequilibrium (LD) measures D'/r^2 values for SNPs in CYP4F2 gene (case-parent design)

CYP4F2										
	rs1272	**rs3093158**	**rs3093193**	**rs3093198**	**rs2016503**	**rs2108622**	**rs2074900**	**rs2074902**	**rs3093145**	**rs3093144**
rs1272	0.24									
rs3093158	0.94/0.45	0.38								
rs3093193	0.90/0.15	0.37/0.05	0.37							
rs3093198	0.86/0.60	0.66/0.27	0.93/0.20	0.28						
rs2016503	0.92/0.56	1.00/0.34	0.86/0.09	0.88/0.41	0.17					
rs2108622	0.94/0.11	1.00/0.24	0.87/0.52	0.96/0.14	1.00/0.08	0.28				
rs2074900	0.84/0.09	0.90/0.21	0.82/0.16	0.88/0.13	0.91/0.07	0.81/0.11	0.29			
rs2074902	1.00/0.07	1.00/0.13	0.93/0.33	1.00/0.08	1.00/0.04	1.00/0.55	0.76/0.05	0.18		
rs3093145	0.80/0.16	0.88/0.37	0.33/0.05	0.45/0.06	0.78/0.10	0.17/0.01	0.85/0.39	0.96/0.16	0.44	
rs3093144	0.74/0.03	0.84/0.07	0.38/0.04	0.15/0.01	0.51/0.01	0.44/0.08	0.97/0.07	1.00/0.04	1.00/0.22	0.15

Off-diagonal elements are estimates of D'/R-squared; Diagonal elements are relative frequencies of allele 2

Table IX-b – Linkage disequilibrium (LD) measures D'/r^2 values for SNPs in ALOX5 gene (case-parent design).

ALOX 5				
	rs3780901	**rs2291427**	**rs10751383**	**rs2115819**
rs3780901	0.32			
rs2291427	0.64/0.35	0.29		
rs10751383	0.61/0.25	0.81/0.38	0.41	
rs2115819	0.23/0.03	0.21/0.02	0.48/0.16	0.50

Off-diagonal elements are estimates of D'/R-squared; Diagonal elements are relative frequencies of allele 2

Table X - Association between CD and SNPs in the CYP4F2 gene. Conditional on parental genotype analysis, N=160 trios.

	Chromosome Location 19pter-p13.11	SNP	Relative Risk	AIC	95% CI	p-value
1.	15850040	rs1272				
		Multiplicative	1.29	546.50	0.87-1.91	0.20
		Recessive	1.14	547.82	0.73-1.79	0.56
		Dominant	3.10	544.11	1.00-5.00	0.04*
2.	15851048	rs3093198				
		Multiplicative	0.95	609.39	0.67-1.35	0.790
		Recessive	0.94	609.42	0.43-1.89	0.853
		Dominant	0.95	609.41	0.62-1.47	0.83
3.	15851431	rs2108622				
		Multiplicative	1.21	651.48	0.86-1.69	0.27
		Recessive	1.08	652.66	0.56-1.98	0.81
		Dominant	1.34	651.07	0.86-2.11	0.20
4.	15852914	rs3093193				
		Multiplicative	0.99	680.96	0.73-1.41	0.933
		Recessive	1.08	680.85	0.66-2.08	0.738
		Dominant	0.93	680.86	0.60-1.44	0.73
5.	15857820	rs2074900				
		Multiplicative	1.03	621.46	0.74-1.43	0.86
		Recessive	0.94	621.46	0.44-1.87	0.86
		Dominant	1.07	621.39	0.71-1.61	0.76
6.	15861166	rs3093158				
		Multiplicative	1.63	635.46	1.17-2.30	3.5 x 10-3*
		Recessive	1.34	641.98	0.89-2.03	0.15
		Dominant	4.02	630.72	1.82-6.5	2.7 x 10-4*
7.	15862852	rs3093145				
		Multiplicative	1.35	704.13	0.99-1.86	0.057
		Recessive	1.11	707.55	0.68-1.80	0.662
		Dominant	1.76	702.10	1.10-2.90	0.017*
8.	15863297	rs3093144				
		Multiplicative	1.32	493.35	0.89-1.97	0.16
		Recessive	1.00	495.29	0.32-2.67	0.99
		Dominant	1.44	492.76	0.92-2.30	0.11
9.	15866185	rs2016503				
		Multiplicative	1.35	448.53	0.86-2.17	0.19
		Recessive	1.29	449.24	0.77-2.16	0.32
		Dominant	2.38	448.78	0.61-4.2	0.23
10.	15869099	rs2074902				
		Multiplicative	1.21	521.92	0.82-1.78	0.33
		Recessive	1.05	522.86	0.40-2.43	0.92
		Dominant	1.30	519.91	0.82-2.10	0.27

Table XI - Association between CD and SNPs in the ALOX gene. Conditional on parental genotype analysis, n=160 trios.

	Chromosome Location 10q11.2 Position	SNP	Relative Risk	AIC	95% CI	p-value
1.	45221095	rs2115819				
		Multiplicative	1.00	710.72	0.73-1.36	0.99
		Recessive	0.91	710.54	0.55-1.44	0.67
		Dominant	1.11	710.54	0.69-1.78	0.68
2.	45237382	rs3780901				
		Multiplicative	0.99	645.22	0.71-1.37	0.93
		Recessive	1.41	644.00	0.76-2.54	0.27
		Dominant	0.82	644.45	0.54-1.27	0.39
3.	45256230	rs2291427				
		Multiplicative	0.96	611.44	0.70-1.35	0.83
		Recessive	1.12	611.38	0.57-2.10	0.74
		Dominant	0.89	610.71	0.60-1.39	0.61
4.	45257998	rs10751383				
		Multiplicative	1.00	711.80	0.74-1.38	0.94
		Recessive	0.94	711.76	0.58-1.51	0.81
		Dominant	1.09	711.68	0.68-1.77	0.72

Table XII - Association between CD and haplotypes of CYP4F2 tag-SNPs in the case-parent design (N=160 trios)

	SNP (Location)											
	RS1272 (15850040)	**RS3093198** (15851048)	**RS2108622** (15851431)	**RS3093193** (15852914)	**RS2074900** (15857820)	**RS3093158** (15861166)	**RS3093145** (15862852)	**RS3093144** (15863297)	**RS2016503** (15866185)	**RS2074902** (15869099)	**FREQUENCY***	**P-VALUE**
H1	G	C	C	G	A	A	A	G	T	T	0.30	0.80
H2	G	C	T	C	G	A	C	G	T	C	0.19	0.26
H3	C	C	C	G	G	G	C	G	C	T	0.15	0.16
H4	G	C	T	C	G	A	A	A	T	T	0.12	0.64
H5	C	T	C	G	G	G	C	G	T	T	0.07	0.61
H6	G	T	C	G	G	A	A	A	T	T	0.06	**0.05***
H7	G	C	C	C	G	G	C	G	T	T	0.05	**0.001***
H8	G	C	C	G	G	G	C	G	T	T	0.05	0.57
H11	G					A	A				0.48	**0.045***
H12	C					G	C				0.22	0.21
H13	G					A	C				0.19	0.31
H14	G					G	C				0.11	**0.004***

*Frequency in the study population

Omnibus test p-value:
All haplotypes= 0.035, 3 SNP haplotype=0.007

Table XIII – parent-of-origin effects analysis examining all ten SNPs in the cyp4f2 gene. Analysis based on N=160 complete trios.

	CHR 19pter-p13.11 Position	SNP	A1:A2	T:U_PAT	CHISQ_ PAT	P_PAT	T:U_MAT	CHISQ_ MAT	P_MAT	Z_POO	P_POO
1.	15850040	rs1272	2:1	20:33	3.19	**0.074**	25:25	0	1	-1.25	0.21
2.	15851048	rs3093198	2:1	27:34	0.80	0.37	35:31	0.24	0.62	-0.99	0.32
3.	15851431	rs2108622	2:1	35:32	0.13	0.71	38:26	2.25	0.13	-0.82	0.41
4.	15852914	rs3093193	2:1	35.5:39.5	0.21	0.64	35.5:30.5	0.38	0.54	-0.76	0.44
5.	15857820	rs2074900	2:1	38.5:35.5	0.12	0.73	31.5:32.5	0.015	0.90	0.33	0.74
6.	15861166	rs3093158	2:1	27.5:47.5	5.33	**0.021**	26.5:41.5	3.31	0.069	-0.28	0.778
7.	15862852	rs3093145	2:1	51:39	1.60	0.20	40:28	2.12	0.145	-0.27	0.785
8.	15863297	rs3093144	2:1	30.5:24.5	0.65	0.41	25.5: 17.5	1.49	0.22	-0.38	0.70
9.	15866185	rs2016503	2:1	13:22	2.31	0.13	18:20	0.105	0.745	-0.88	0.382
10.	15869099	rs2074902	2:1	27.5:25.5	0.075	0.78	29.5:20.5	1.62	0.20	-0.72	0.47

Table XIV. Demographic and clinical characteristics of the CD Patients (N=225) and controls (N=330) in the case-control design.

CASES **N=225**		**CONTROLS** **N=330**		**P-value**
Age at diagnosis		**Age at diagnosis**		
Mean (SD)	12.4 (3.4)	Mean (SD)	10.6 (7.2)	**0.005**
Study site (%)		**Study site (%)**		
Montreal	**173 (76.9)**	**Montreal**	**282 (84.4)**	
Ottawa	**52 (23.1)**	**Ottawa**	**48 (14.6)**	**0.01**
Gender (%)		**Gender (%)**		
Females	87 (38.7)	Females	131 (39.7)	**0.81**
Males	138 (61.3)	Males	199 (60.3)	
		Type		
		Orthopedic + siblings	**225 (68.1)**	
		Birth cohort	**20 (6.1)**	
		Healthy adults	**85 (25.8)**	

CD cases - clinical characteristics (at diagnosis)	
Disease location* N=225	**N (%)**
Ileum with or without upper digestive tract, L1+/-L4	40 (17.8)
Colon with or without upper digestive tract, L2+/-L4	59 (26.2)
Ileum and colon with or without upper digestive tract, L3 +/-L4	126 (56.0)
Disease behaviour* N=224 (%) (The disease behaviour of 1 patient could not be appropriately identified)	**N (%)**
Inflammatory B1 +/- p‡	190 (84.8)
Stricturing B2 +/- p‡	22 (9.8)
Perforating B3 +/- p‡	12 (5.4)

*according to Montreal classification

‡ perianal modifier

Table XV-a – allelic and genotype associations between the rs1272 dna variant in the cyp4f2 gene and cd. Analysis based on the case-control study [a] (n=225 cd cases; m=330 controls).

Chromosome Location 19pter-p13.11 Position	SNP	Cases N=225 (%)	Controls M=330 (%)	Odds ratio (95% CI)	P-value	AIC[b]
CYP4F2 rs1272 (15850040)	G	350 (77.8)	514 (77.9)	1.00 (0.74-1.35)	0.48*	N/A
	C	100 (22.2)	146 (22.10)			
	CC	10 (4.4)	23 (7.0)	Reference	0.13*	752.63
	GC	80 (35.6)	100 (30.3)	1.84 (0.83-4.1)	0.13*	
	GG	135 (60.0)	207 (62.7)	1.50 (0.69-3.2)	0.30	
	Recessive Model			0.89 (0.63-1.26)	0.52	752.98
	Dominant Model			1.61 (0.75-3.45)	0.22	751.83

[a] The actual number of cases and controls per SNP can vary according to the genotyping call rates; [b] Akaike Information Criteria; * 1-sided p-value

Table XV-b– allelic and genotype associations between the rs3093158 dna variant in the cyp4f2 gene and cd. Analysis based on the case-control study[a]
(N=225 cases; m=330 controls)

Chromosome Location 19pter-p13.11 Position	SNP	Cases N=225 (%)	Controls M=330 (%)	Odds ratio (95% CI)	P-value	AIC[b]
CYP4F2 rs3093158 (15861166)	C	307 (68.9)	430 (65.2)	1.18 (0.91-1.54)	0.10*	N/A
	T	139 (31.2)	230 (34.8)			
	TT	18 (8.1)	48 (14.5)	Reference	0.03*	745.89
	TC	103 (46.2)	134 (40.6)	2.05 (1.12 - 3.73)	0.019	
	CC	102 (45.7)	148 (44.9)	1.84 (1.01 - 3.34)	0.046	
	Recessive Model			1.04 (0.74 – 1.46)	0.84	749.74
	Dominant Model			1.94 (1.09 – 3.43)	0.023	744.24

[a] The actual number of cases and controls per SNP can vary according to the genotyping call rates; [b] Akaike Information Criteria; * 1-sided p-value

Table XV-c– allelic and genotype associations between the rs3053145 dna variant in the cyp4f2 gene and cd. Analysis based on the case-control study [a]
(n=225 cases; m=330 controls).

Chromosome Location 19pter-p13.11 Position	SNP	Cases N=225 (%)	Controls M=330 (%)	Odds ratio (95% CI)	P-value	AIC[b]
CYP4F2 rs3053145 (15862852)	G	239 (53.1)	356 (54.1)	1.04 (0.81-1.33)	0.74	N/A
	A	211 (46.9)	302 (45.9)			
	GG	62 (27.6)	95 (28.9)	Reference	0.94	754.24
	GA	115 (51.1)	166 (50.5)	1.06 (0.71-1.58)	0.77	
	AA	48 (21.3)	68 (20.7)	1.08 (0.66-1.76)	0.75	
	Recessive Model			1.04 (0.69-1.58)	0.85	752.33
	Dominant Model			1.07 (0.73-1.56)	0.73	752.25

[a]The actual number of cases and controls per SNP can vary according to the genotyping call rates; [b]Akaike Information Criteria; * 1-sided p-value

Table XVI - Association between CYP4F2 haplotypes and risk for CD in children. Analysis based on the combined case-control data (n=225 cases, m=330 controls)

	CYP4F2 SNP (LOCATION)					
	RS1272 (15850040)	**RS3093158** (15861166)	**RS3093145** (15862852)	**MAJOR ALLELE FREQUENCY IN CASES**	**MAJOR ALLELE FREQUENCY IN CONTROLS**	**P-VALUE**
H1	C	A	A	0.48	0.46	0.42
H2	G	G	C	0.21	0.22	0.85
H3	C	A	C	0.21	0.19	0.30
H4	C	G	C	0.09	0.13	0.02*

* P <0.05

Table XVII – Interaction between CYP4F2 SNP rs3093158 and gender in Canadian children with CD. Finding based on the logistic regression analysis applied to all cases (N=385) and the controls (M=330).

SNP	**Model**	**Odds ratio (95%CI)**		**Interaction Odds ratio (p-value)**
		Females	**Males**	
CYP4F2 rs3093158	Dominant†	2.31 (1.03-5.20)	2.33 (1.21-4.48)	1.00 (0.98)
	Additive↑	1.11 (0.78-1.60)	1.31 (0.97-1.74)	1.17 (0.51)

† Based on coding the homozygous common genotype=0 and the heterozygous and homozygous rare genotypes as 1; * $P \leq 0.05$.

Table XVIII - Location of the tag-SNPs in CYP4F2 and ALOX5 genes and their frequencies in the unrelated controls (case-control data).

Gene/ Chromosomal location		SNP	Polymorphism	Minor allele frequency in the unrelated controls
CYP4F2 Chromosome 19pter-p13.11				
1.	15850040	rs1272	G/C	22.1%
6.	15861166	rs3093158	C/T	34.8%
7.	15862852	rs3093145	G/A	45.9%
ALOX5 Chromosome 10q11.2				
1.	45221095	rs2115819	A/G	44.0%
2.	45237382	rs3780901	C/T	34.8%
3.	45256230	rs2291427	A/G	32.4%
4.	45257998	rs10751383	A/C	46.4%
NOD2/CARD15 Chromosome 16q21				
1.	49303427	rs2066844 (SNP8)	C/T	8.5%
2.	49314041	rs2066845 (SNP12)	G/C	1.8%
3.	49321282^49321283	rs5743293 (SNP13)	insC	4.5%

Table XIX - Interactions between CYP4F2, ALOX5 and NOD2 genes. Findings based on applying the multifactor dimensionality Reduction* to the combined case population (n=385) versus unrelated controls (m=330).

Model	**Training accuracy**	**Testing Accuracy**	**Cross-validation**	**P-value****
rs3093158 (CYP4F2), NOD2/CARD15	0.5989	0.5806	8/10	<0.005
rs3093158 (CYP4F2), NOD2/CARD15, **rs2115819 (ALOX5)**	0.6196	0.5795	9/10	<0.007

* Subjects with 1 or more variants of the three NOD2/CARD15 SNPs (SNP8, SNP12 and SNP13) were designated as NOD2+, and those with none as NOD2-

** Based on 1000 permutations

Table XX - Case-based analysis of association between CYP4F2 SNP rs3093158 with CD disease location and behaviour in Canadian children. Findings based on the combined set of cases (n=385).

SNP	Variant	Disease location		Odds ratio (95% CI)	P-value
		Ileal*	**Colonic**		
CYP4F2 rs3093158	C	160 (29.9)	63 (31.5)	1.07 (0.74-1.54)	0.69
	T	374 (70.0)	137 (68.5)		
	CC	9 (9.0)	16 (6.0)		0.57
	CT	45 (45.0)	128 (47.9)		
	TT	46 (46.0)	123 (46.1)		
	Disease behaviour				
	Variant	**Inflammatory**	**Non-inflammatory****	**Odds ratio (95% CI)**	**P-value**
	C	441 (69.8)	68 (68.0)	1.09 (0.67-1.74)	0.72
	T	191 (0.30)	32 (32.0)		
	CC	144 (45.6)	24 (48.0)		0.23
	CT	153 (48.4)	20 (40.0)		
	TT	19 (6.0)	6 (12.0)		

* includes ileal disease with or without colonic involvement

** Includes structuring and/or penetrating disease

Table XXI - Case-control analysis of association between CYP4F2 SNP rs3093158 with CD disease location and behaviour in Canadian children. Findings based on the combined set of cases (n-385) versus unrelated controls (m=330).

SNP	Controls (%)	Ileal*		Colonic	
		Odds ratio (95% CI)	**P-value**	**Odds ratio (95% CI)**	**P-value**
rs3093158 (C/T)	34.8%	1.25 (0.97-1.61)	0.07	1.16 (0.82-1.66)	0.38
		Inflammatory		Non-inflammatory**	
		1.23 (0.97-1.57)	0.08	1.14 (0.71-1.84)	0.57

* includes ileal disease with or without colonic involvement

** Includes stricturing and/or penetrating disease

Section E. Discussion

Our analyses based on the exploratory phase (case-parent design) implicated three genetic variants in the CYP4F2 gene in the etiopathogenesis of CD in children. Using the case-control design in Phase 2, association between 1 tag-SNP in the CYP4F2 gene (rs3093158) and CD was successfully replicated. Furthermore, case-control analysis revealed that interactions between the CYP4F2, ALOX5 and NOD2/CARD15 genes could play an important role in disease susceptibility. There was no evidence for associations between CYP4F2 and gender or CYP4F2 and specific disease phenotypes. No parent-of-origin effects were obvious when studying the transmission of either the CYP4F2 and/or ALOX5 gene variants to affected children.

CD is a multi-factorial disease with a strong genetic component. Few disorders in clinical medicine are associated with as much chronic morbidity as CD. The disorder is characterised by transmural inflammation that could affect any part of the GIT, and the disease relapses and remits throughout its course. For many patients, CD imposes a substantial personal burden, with unpredictable fluctuating symptoms, time off work, need for expensive drugs, or surgery and multidisciplinary care. This becomes more evident in early-onset of CD, which is marked by growth failure, higher disease severity and frequency of complications when compared to adult onset of disease [17]. With a prevalence of about 0.5% in Canada [1], CD also places a major burden on public health-care resources.

In spite of a wealth of information that suggests that alternative pathways involved in the innate or adaptive immunity exert feedback regulation resulting in aberrant and chronic inflammation and disease, the exact mechanisms underlying the inflammatory process in the GIT are unknown. Considering chronic inflammation as a key point in disease development, we postulated that genetic variants that code for key enzymes involved in the metabolism of pro-inflammatory mediators, in particular eicosanoids, could play an important role in early-onset disease susceptibility. More specifically, we targeted 2 candidate genes, the ALOX5 and the CYP4F2 genes that are intricately linked to the production of important inflammatory mediators, such as prostaglandins and leukotrienes that are known to trigger and maintain the inflammatory process.

To our knowledge this is the first study that targeted mechanistic pathways of inflammation in order to elucidate the complex pathogenesis of CD. The focus of this study was the 5-LO pathway, a widely studied pathway of inflammation. Via the 5-LO pathway, arachidonic acid (AA) is transformed to biologically active pro-inflammatory mediators. The eicosanoid family of mediators includes prostaglandins, thromboxanes, and leukotrienes. Upon activation, the immediate product of 5-LO pathway is LTA_4 (leukotriene A_4), which is enzymatically converted into LTB_4 (leukotriene B_4) by LTA_4 hydrolase. Biologically active LTB_4 is metabolized by omega-oxidation carried out by specific cytochrome P450s (CYP4F), ultimately yielding the various eicosanoid metabolites. These metabolites have well-established roles in many pathological processes, including recruitment and activation of neutrophils, thrombosis, inflammation and immunosuppression [221, 224, 289, 290], and disease-states (e.g. asthma, allergic rhinitis, idiopathic pulmonary fibrosis, atherosclerosis, atopic dermatitis, acne, ischemia-related organ injury, and IBD) [291].

Defects in expression of participating enzymes or genetic variation of their encoding genes may be involved in dysregulation of this pathway and susceptibility to inflammation. Prior evidence suggested a possible link between variants in these genes and inflammation. For instance, arachidonate 5-lipoxygenase is the key enzyme in leukotriene biosynthesis and catalyzes the initial steps in the conversion of AA to biologically active leukotrienes. This enzyme is encoded by the 5-LO gene (i.e. ALOX5 gene) that has been linked to atherosclerosis in mouse models and in humans [238, 239]. Recently, a tandem-repeat polymorphism in the promoter region of the ALOX5 gene was shown to be associated with carotid intimal-medial thickness, a measure of atherosclerosis [242, 292]. This polymorphism was also related with ALOX5 mRNA expression and leukotriene production, and with the response to anti-asthma treatments [240, 241].

Another candidate gene of interest is CYP4F2, which is located in a region shown previously to be linked to IBD (i.e. 19p). A recent study by Curley et al. revealed that IBD and coeliac disease, another chronic inflammatory disorder of the gastrointestinal system, share two linkage regions: on chromosome 5q31 (CELIAC2 and IBD5), and 19p13 (CELIAC4 and IBD6). The same study revealed that the CYP4F2 gene showed a small effect on familial clustering of this disease, suggesting a putative role of this gene in disease pathogenesis [252]. Other studies investigating potential links between CYP4F2 and inflammation are presently lacking.

As the onset and maintenance of inflammation is likely to be dependent upon AA-derived prostaglandins and eicosanoids, genetic mutations in genes that encode for enzymes that are implicated in this pathway can tip the balance of this metabolic process. Thus, we postulated that genes that code for enzymes within the pathway are likely to play important roles in the determination of susceptibility for inflammation and indirectly in the susceptibility for CD. Thus, individuals' who are genetically susceptible by way of harbouring genes that negatively influence this pro-inflammatory metabolic pathway could be more likely to be at risk for CD.

The main objective of this study was to examine whether the ALOX5 gene, which encodes for the 5-LO enzyme – a key enzyme in the 5-LO pathway, and in the CYP4F2 gene that encodes the cytochrome p450 enzyme, and which is specifically implicated in the omega-oxidation of the biologically active LTB_4, could be relevant to the development and perpetuation of inflammation in the gut. The CYP4F2 gene is located in chromosome region 19p13.11. Three genome scans and a meta-analysis of GWAS have reported significant linkages to this region in different populations. Rioux et al (2000) using affected sib-pairs reported peak LOD scores of 4.6 for IBD and 3.0 for CD in a Canadian population [75]. In their recent meta-analyses of 10 IBD genome scans among affected relatives, van Heel et al (2004) reported significant linkage of CD to chromosome 19 [248]. Low et al (2004) carried out a linkage scan among UK Caucasians (affected sib-pairs) and confirmed linkage of CD to the 19p13.2 region (peak multi-point linkage score of 1.59) [249]. Finally, a meta-analysis of GWAS complemented by an independent replication study totalling 5,555 CD cases and 6,638 controls, confirmed the association between CD and the 19p13 region (critical region 1.05-1.15), i.e. RR=1.26 [254]. The location of the CYP4F2 gene in the region of significant GWAS signals adds further support to our observations that it may be an important candidate gene for CD.

The exact mechanisms whereby variation in the CYP4F2 gene can predispose to CD can only be speculated on. As the CYP4F2 gene influences neutrophil migration, variants in the gene may alter this process and lead to a dysregulated response to inflammatory stimuli whereby there is excess neutrophil trafficking and chemotaxis to the intestinal epithelium. The latter could lead to a cascade of events that includes stimulations of pro-inflammatory cytokines that can sustain the inflammatory process. It has been reported that activated neutrophils have an effect on epithelial tight junctions that results in enhanced permeability [293] in the intestine, which could facilitate the influx of various commensals from the lumen leading to the recruitment of even more phagocytic neutrophils and thus to a perpetuation of barrier impairment and hence inflammation. When this balance in mucosal immune-system responsiveness is tilted towards an exaggerated or uncontrolled reaction against the commensal flora, CD may result. This is in line with a previous hypothesis that states that CD may be caused by genetic defects that produce exaggerated innate responses to the flora leading to excessive inflammation [294]. Concomitantly, genetic variants in the CYP4F2 gene may directly influence neutrophil migration, a feature usually associated with the innate immune response and thus create an environment in the gut mucosa that contributes to CD pathogenesis. There are a number of congenital conditions such as Glycogen Storage Disease type 1b [295], Chediak-Higashi syndrome [296] and Leukocyte Adhesion Deficiency syndrome [297] that are associated with CD and all of these conditions have poorly functioning neutrophils. Furthermore, the use of a stimulator of neutrophils showed evidence suggestive of decreased CD activity [298].

Epidemiologic evidence suggests that in addition to genetic variation *per se*, epigenetic factors may play an important role in the pathogenesis of pediatric CD. In particular, findings such as: (i) the maternal effect in transmission of CD from affected parent to offspring; (ii) parental differences in the degree of genetic anticipation; and, (iii) discordance of monozygotic twins affected by CD, have provided strong support to this hypothesis. Because such parent-of-origin effects may be of pathogenic importance in CD, one of our secondary objectives was to explore if there were any differences in the transmission of the risk alleles from parents of one sex to children. In our case-population, no such preferential transmission was observed for the CYP4F2 gene.

Assuming a key role played by the 5-LO enzymes in the pathogenesis of inflammation, it can be thought that alterations of several enzymes within this pathway would be more predictive of the outcome, i.e. CD, than either of them alone. Concomitantly, existing evidence underlines the importance of microbial-mucosal interaction in the pathogenesis of CD and the modulating role of NOD2/CARD15 gene in this process [200, 201]. We thus postulated that interactions between these genes (i.e. ALOX5, CYP4F2 and NOD2/CARD15) could modulate the inflammatory process and a combination of variants in these genes may impact disease. To test this hypothesis, we analyzed whether interactions between the 4 SNPs in the ALOX5 gene, 3 SNPs in the CYP4F2 gene and carriage of any of the 3 SNPs (SNP8, SNP12 and SNP13) in the NOD2/CARD15 gene were associated with CD. To study this interaction, we applied MDR methods to the case-control data. A three-SNP model that included the CYP4F2 SNP rs3093158, the ALOX5 SNP rs2115819 and the NOD2/CARD15 gene was significantly associated with disease, suggesting that in combination these genes may have an impact on disease susceptibility. It is of interest to note that the ALOX5 SNPs were not independently associated with CD in our study. Nonetheless in the presence of variants in both the NOD2/CARD15 and CYP4F2 genes, it was found to impact disease. These observations are in line with the biologic process leading to CD wherein an interaction between innate immunity (CARD15) and acquired immunity (ALOX5, CYP4F2) is implicated.

Rather than a unique disorder, CD is regarded as a collection of various inflammatory-driven disorders of the GIT. Biological discrimination between various subgroups of CD patients led to a proposal for a phenotypic classification, which was taken up and further developed and evaluated by an international working group, leading first to the Vienna classification [32], followed by the Montreal classification [33] – **Table I**. Accordingly, CD is separated into three categories: age at diagnosis, location, and behaviour. Several teams have now tested these classifications in their databases and found that they are feasible and reproducible [30, 299, 300]. It is evident that variations in the human genome contribute to disease susceptibility and also shape the clinical phenotype. For instance, in addition to conferring risk to disease susceptibility, variation in the NOD2/CARD15 gene also it relates to ileal disease location, possibly through its selective expression in Paneth cells and regulation of antibacterial host defence [144, 301]. This state-of-knowledge prompted us to seek out any associations between the identified risk allele in the CYP4F2 gene (i.e. rs3093158) and disease location or behaviour. No significant impact of rs3093158 on disease location and/or behaviour was observed. However, these analyses were limited by power and need to be examined in larger samples.

Another interesting phenotypic particularity of early-onset CD is a preponderance of male versus female cases. In the context of our study, investigating the interaction between gender and the CYP4F2 risk allele was considered to be of interest in providing further insight into the pathogenesis of the disease. No significant interaction between gender and rs3093158 was however observed. Given the relatively low power of a stratified gender-based analysis, larger studies would be required to detect gender-specific associations between CYP4F2 and CD, if any.

Strengths and limitations of the study

We carried out a two-phase candidate gene study, employing the case-parent design in Phase 1 and the case-control design in Phase 2, to examine the impact of genes within the 5-LO pathway and risk for CD. To underline the significance of our results, the combined p-value is highly suggestive of association with one SNP rs3093158 and CD (combined corrected $p_{multiplicative}$=0.015, combined corrected $p_{dominant}$=4.53 x 10^{-4}). In our context, we used a strict definition of a replication study, meaning that we explored a genetic association in the replication phase with (i) the same phenotype, (ii) only using the SNPs that were found to be significantly associated with CD in the first stage of our study, and (iii) the same direction of effects [302]. In general replication studies require investigations in populations that are independent of the original study. For our study, the replication case-population included cases that were initially recruited as part of the Phase 1 study, but for whom one or both parents were not available (i.e. incomplete trios, ~23 cases). So in some ways the replication population was not independently selected. Furthermore, the replication population of cases was a subset of a larger cohort (as the replication stage was planned subsequent to the findings in the Phase 1 of the study). However, as the selection factors that might have influenced any disparities between the two case-populations were similar and not under the influence of the exposure under study (i.e. genotypes), we believe that the possibility of them influencing the validity of the findings in a significant way is unlikely. In addition, other biases if any are likely to be limited as (i) the clinical characteristics at diagnosis of the patient population was by-and-large similar – **Table XXII**; (ii) the allele frequencies of the three CYP4F2 SNPs (i.e. rs3093158, rs1272, and rs3093145) were by-and-large similar within the pseudo-controls (case-parent design) and the unrelated controls (case-control study), indicating that both samples were generated from the same source population; and (iii) the major allele of rs3093158 seemed to be risk-conferring allele in both the exploratory and replication stages.

By using the proposed case-parent approach in the Phase 1 study, the problem of population stratification inherent to the case-control design is completely eliminated by the use of non-transmitted alleles as controls. Also in using this approach, genetic associations are robust to confounding from non-genetic exposures under the principle of Mendelian randomization [303] whereby alleles are transmitted from parents to offspring randomly, simulating a randomized trial and thus ensuring that non-genetic factors are randomly distributed with genotype groups in the population. However, non-genetic exposures that are known risk factors for the disease of interest can confound genetic associations, if: [1] the genotype influences the exposure (e.g. possessing specific alleles in the alcohol dehydrogenase gene (ADH) prevents individuals from consuming alcohol, resulting in a negative correlation between the genotype and alcohol consumption. A study that examines association between the gene and heart disease would need to consider alcohol consumption as a potential confounder). [2] If the gene being studied, is in LD (or associated) with another gene that influences the exposure (in the example above, if ADH did not directly influence alcohol consumption but was associated with the aldehyde dehydrogenase (ALDH) gene that influenced alcohol consumption, alcohol consumption would be potential confounder).

As previously mentioned, one of the main challenges, when implementing the case-control design in a pediatric population, is the selection of healthy population-based controls. By using controls from three different sources (i.e. healthy children visiting the orthopaedic departments, healthy children from the general population, and healthy French-Canadian adults), we ensured that the control population was similar to the source population. Still, our case-control design could have been affected by potential confounding due to population stratification. In order to control for this, only Caucasian subjects were included. Nonetheless, there were differences in the magnitude of the associations between the risk alleles and CD in the case-parent study versus the case-control study, some of which could be attributed to subtle population stratification.

We included both newly-diagnosed and prevalent cases in both our exploratory and replication stages. The possibility of selection bias influencing the results was considered. It is likely that some patients who had CD were missed, either because they were transferred to other hospitals (having crossed the age limit of 20 years) or died due to the disease. In respect to the former, the majority of our cases were diagnosed after 1988 and they are currently under follow-up in a pediatric hospital. Intentionally, we did not include in this study cases that were diagnosed before 1988 and who could have left the hospital. We also believe that it is unlikely that the distribution of alleles in the cases we possibly missed versus the ones we studied to be different. In respect to the latter, CD is not a fatal disease and therefore it is improbable that we missed cases because they died before being recruited. Another source of selection bias that might also affect the external validity of our study is the level of disease severity. In the beginning, mild cases of CD might get treated by family physicians. CD is a long-standing condition with periods of relapses and remittance and most of the patients (if not all) will be eventually referred to a gastroenterologist. If the disease has a spontaneous remittance, it probably means that it is not CD and the patient was misdiagnosed.

The inclusion of incident/prevalent cases could influence the examination of genotype-phenotype relationships. This is particularly true for diseases such as CD where the disease phenotype evolves over time. Thus the timing of ascertainment of the phenotype is important. Selection bias would be possible if for some cases phenotypes at diagnosis were included whereas for others phenotypes during disease course were considered. In order to avoid this we focussed on disease phenotype at diagnosis for all cases. At the same time with regards to susceptibility for a particular phenotype, it is the phenotype at diagnosis that is likely to be relevant for prevention/diagnosis.

Information bias, i.e. genotyping accuracy, needs to be considered for all genetic studies including ours. In using the GenomeLab™ Sequenom Genotyping platform, we ensured that our genotyping results were highly accurate (i.e. 96% success rate, 85% call rate, and error rate less than 0.5%). In addition, 20% of the samples were randomly re-genotyped for all the SNPs to ensure reproducibility. The reproducibility rate was >95%. As genotyping was carried out blinded to the case/parent/control status of the subject, misclassification bias if any, was further likely to be minimal.

With regards to confounding, environmental factors by-and-large are unlikely to confound genetic associations. The only environmental factor that has been consistently associated with adult CD is smoking. There is limited evidence for smoking as a risk factor for pediatric-onset CD. Smoking is rare prior to age 10 and among those between age 11-15, the prevalence of smoking is approximately 15% (ref). CD is most common between ages 10-15 and smoking may act as a trigger for disease. However, collecting information on smoking from adolescents is likely to be unreliable. On the other hand however, children exposed to passive smoke may be at higher risk. We have however shown that maternal smoking was not associated with pediatric CD [11], indicating limited support for smoking as a risk factor for pediatric-CD. Although environmental risk factors are unlikely to be major confounders in our study, it is possible that other genetic confounders might be responsible for the observed effect. For instance, genes that are in LD with CYP4F2, such as the neighbouring CYP4F3 gene, may confound these associations. However, none of the CYP4F2 genes we studied (in particular SNP rs3093158 that was found to be associated with CD), are in LD with SNPs in the CYP4F3 gene suggesting that confounding due to LD may also be limited.

We compared our findings of association with CYP4F2 gene with that of the WTCCC study [254] and that of Curley et al [252]. In the former, SNPs in and around the CYP4F2 were not associated with CD. The WTCCC study however was exclusively based on adult CD. There is growing evidence that genetic susceptibility may be different in early- versus adult-onset disease [18]. Similarly, it is not clear whether the genome-wide SNPs studied by the WTCCC provided adequate coverage of SNPs within the CYP4F2 region and also whether their study had sufficient power to capture modest associations. This is highlighted by findings from the meta-analyes of GWAS [254] that reported significant associations with SNPs within the region of the CYP4F2 gene.

The Curley et al study (2006) revealed that 2 tag SNPs in the CYP4F2 gene were associated with risk for celiac disease in an adult Dutch celiac disease cohort. One of these SNPs, rs3093156, is in strong LD with rs3093145 (LD = 1.0), a SNP that was significantly associated with CD in our exploratory phase. Given that both celiac disease and CD are characterized by gut inflammation and therefore potentially share some genetic susceptibility, these observations imply that CYP4F2 may play an important role in gut inflammation that is characteristic of CD. As associations with SNP rs3093145 were not replicated in our case-control study, certainly larger studies would be required to assess whether SNP rs3093145 is also associated with pediatric CD.

Gene-gene interaction is a fundamental component of the genetic architecture of complex traits such as disease susceptibility. Interaction has been recognized for many years and has been described essentially from two different perspectives, biological and statistical [154]. Biological interaction, as defined by William Bateson in 1909, results from physical interactions among bio-molecules in gene regulatory networks and biochemical pathways at the cellular level in an individual. Statistical interaction, as defined by Fisher shortly thereafter (1918) in «The Correlation Between Relatives on the Supposition of Mendelian Inheritance», is deviation from additivity in a linear mathematical model that describes the relationship between multilocus genotypes and phenotype variation at the population level [304]. Gene-gene interaction, along with other phenomena such as locus heterogeneity, phenocopy, and gene–environment interaction, is a major source of complexity in the mapping relationship between genotype and phenotype. As such, a need for research strategies that embrace, rather than ignore, this complexity has been described. In the current study, we addressed this necessity by implementing a recently developed non-parametric and genetic model-free data mining algorithm called MDR [283, 284]. With MDR, multilocus genotypes are pooled into high-risk and low-risk groups, effectively reducing the dimensionality of the genotype predictors to one dimension. The power of this strategy is the ability to analyze different selection methods and different construction algorithms leading to knowledge discovery from patterns of results across different strategies. The greatest challenge however rests with the interpretation of these models. Another important concern with conducting these analyses with a method such as MDR is that there may be a certain amount of important information potentially lost by limiting results to one best model. An interesting future direction would be to develop hypothesis testing methods that are able to identify a best set of statistically significant MDR models rather than a single best model.

In this study we examined associations with 10 tag-SNPs in the CYP4F2 gene. Two of the tag-SNPs genotyped in the CYP4F2 gene were coding SNPs (rs2108622 and rs2074900). However, associations between them and CD were not evident. In contrast associations were noted with SNPs either in the intronic regions or in the 3' utr regions of the gene. As these latter SNPs were in low LD (r^2<0.80) with the coding tag-SNPs (else they would not be selected as tag-SNPs), it would be important to further assess putative risk-conferring SNPs in the gene.

In this study, we were able to capture most of the newly diagnosed CD cases that presented at either one of the three major tertiary hospitals. Similarly, most cases diagnosed since 1995 and undergoing treatment were captured from the main study hospital in Montreal. Furthermore, the phenotypic distribution in both our case-parent and case-control setting was comparable to the published distribution of phenotypes in children with CD worldwide. Taken together, our findings could therefore by-and-large be generalised to the pediatric Caucasian CD population.

Public health relevance

The early-onset form of CD is of important concern. Firstly, it is becoming more and more apparent that CD incidence and prevalence among children and young adults is on the rise in Canada with rates approaching those in adults. Secondly, children with CD experience certain features unique to pediatric CD, more disease-associated morbidity and the long-term complications of this disease impact greatly on their physical and psychological well-being. Thirdly, it is not known whether potential risk factors contributing to the adult-onset CD similarly predisposes to the early-onset form of disease. Last, to date, the early-onset form of this disease has been insufficiently investigated. Research in the pathogenesis of pediatric CD has thus become of particular public health interest. Studying genetic susceptibility to CD in children is of particular relevance to the further understanding of the pathogenesis of this disease and implicitly to the pursuit of finding ways to prevent the development of this disease.

Conclusions and future directions

We observed that genes involved in the 5-LO metabolic pathway may be important to the pathogenesis of CD. It would be important to further examine these findings in a larger cohort of patients preferably in different geographical and ethnic populations. In parallel, investigating the presence of other functional SNPs in the gene by deep re-sequencing would be relevant. A more detailed assessment of the contribution of CYP4F2 variants to neutrophil recruitment and intestinal permeability that can be related to the development of inflammation associated with CD will be necessary. Fundamental studies to examine the expression of CYP4F2 in CD patients would be interesting. As the 5-LO inflammatory pathway is common to various other related disorders examination of associations of these diseases with variants in the CYP4F2 would be worthwhile. Of relevance would be also to identify other candidate genes in the 5-LO pathway and study their association (either independent or interactions) with CD. Our study findings indirectly suggest that imbalances in dietary fatty acids could impact the pathogenesis of CD. Future studies investigating interactions between dietary consumption of fatty acids, 5-LO pathways genes and risk for CD could provide insights for preventive or therapeutic interventions.

Table XXII - Comparison between the distribution of clinical characteristics at diagnosis in the 2 case-populations

(i.e. case-parent and case-control designs) versus the evidence from the published literature.

Clinical Characteristics at diagnosis	Case-Parent Design (%)	Case-Control Design (%)	Published data for pediatric CD* (%)
Disease Location			
Ileum with or without upper digestive tract	13.7	17.8	5.9
Colon with or without upper digestive tract	28.7	26.2	36.3
Ileum and colon with(out) upper digestive tract	57.5	56	50.5
Disease Behaviour			
Inflammatory	88	84.8	78.8
Stricturing	8.1	9.8	7.7
Perforating	3.1	5.4	8.4

* after (Van Limbergen, Russell et al. 2008)

References

1. Bernstein, C.N., et al., *The epidemiology of inflammatory bowel disease in Canada: a population-based study.* Am J Gastroenterol, 2006. **101**(7): p. 1559-68.
2. Crohn, B.B., L. Ginzburg, and G.D. Oppenheimer, *Landmark article Oct 15, 1932. Regional ileitis. A pathological and clinical entity. By Burril B. Crohn, Leon Ginzburg, and Gordon D. Oppenheimer.* Jama, 1984. **251**(1): p. 73-9.
3. Sands, B.E., *Inflammatory bowel disease: past, present, and future.* J Gastroenterol, 2007. **42**(1): p. 16-25.
4. Farrokhyar, F., E.T. Swarbrick, and E.J. Irvine, *A critical review of epidemiological studies in inflammatory bowel disease.* Scand J Gastroenterol, 2001. **36**(1): p. 2-15.
5. Sonnenberg, A., D.J. McCarty, and S.J. Jacobsen, *Geographic variation of inflammatory bowel disease within the United States.* Gastroenterology, 1991. **100**(1): p. 143-9.
6. Shivananda, S., et al., *Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD).* Gut, 1996. **39**(5): p. 690-7.
7. Yang, S.K., E.V. Loftus, Jr., and W.J. Sandborn, *Epidemiology of inflammatory bowel disease in Asia.* Inflamm Bowel Dis, 2001. **7**(3): p. 260-70.
8. Sandler, R.S. and A.L. Golden, *Epidemiology of Crohn's Disease.* J Clin Gastroenterol, 1986. **8**(2): p. 160-5.
9. Bernstein, C.N., et al., *Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study.* Am J Epidemiol, 1999. **149**(10): p. 916-24.
10. Timmer, A., *Environmental influences on inflammatory bowel disease manifestations. Lessons from epidemiology.* Dig Dis, 2003. **21**(2): p. 91-104.
11. Amre, D.K., et al., *Investigating the hygiene hypothesis as a risk factor in pediatric onset Crohn's disease: a case-control study.* Am J Gastroenterol, 2006. **101**(5): p. 1005-11.
12. Bonen, D.K. and J.H. Cho, *The genetics of inflammatory bowel disease.* Gastroenterology, 2003. **124**(2): p. 521-36.

13. Langholz, E., et al., *Incidence and prevalence of ulcerative colitis in Copenhagen county from 1962 to 1987.* Scand J Gastroenterol, 1991. **26**(12): p. 1247-56.

14. Kugathasan, S., et al., *Epidemiologic and clinical characteristics of children with newly diagnosed inflammatory bowel disease in Wisconsin: a statewide population-based study.* J Pediatr, 2003. **143**(4): p. 525-31.

15. Sawczenko, A., et al., *Prospective survey of childhood inflammatory bowel disease in the British Isles.* Lancet, 2001. **357**(9262): p. 1093-4.

16. Griffiths, A.M., *Specificities of inflammatory bowel disease in childhood.* Best Pract Res Clin Gastroenterol, 2004. **18**(3): p. 509-23.

17. Silbermintz, A. and J. Markowitz, *Inflammatory bowel diseases.* Pediatr Ann, 2006. **35**(4): p. 268-74.

18. Van Limbergen, J., et al., *Definition of Phenotypic Characteristics of Childhood-Onset Inflammatory Bowel Disease.* Gastroenterology, 2008.

19. Mamula, P., et al., *Inflammatory bowel disease in children 5 years of age and younger.* Am J Gastroenterol, 2002. **97**(8): p. 2005-10.

20. Sawczenko, A. and B.K. Sandhu, *Presenting features of inflammatory bowel disease in Great Britain and Ireland.* Arch Dis Child, 2003. **88**(11): p. 995-1000.

21. Heyman, M.B., et al., *Children with early-onset inflammatory bowel disease (IBD): analysis of a pediatric IBD consortium registry.* J Pediatr, 2005. **146**(1): p. 35-40.

22. Levine, A., et al., *TNF promoter polymorphisms and modulation of growth retardation and disease severity in pediatric Crohn's disease.* Am J Gastroenterol, 2005. **100**(7): p. 1598-604.

23. Meinzer, U., et al., *Ileal involvement is age dependent in pediatric Crohn's disease.* Inflamm Bowel Dis, 2005. **11**(7): p. 639-44.

24. Paul, T., et al., *Distinct phenotype of early childhood inflammatory bowel disease.* J Clin Gastroenterol, 2006. **40**(7): p. 583-6.

25. Levine, A., et al., *Pediatric onset Crohn's colitis is characterized by genotype-dependent age-related susceptibility.* Inflamm Bowel Dis, 2007. **13**(12): p. 1509-15.

26. Brown, M.O., *Inflammatory bowel disease.* Prim Care, 1999. **26**(1): p. 141-70.

27. Kanof, M.E., A.M. Lake, and T.M. Bayless, *Decreased height velocity in children and adolescents before the diagnosis of Crohn's disease.* Gastroenterology, 1988. **95**(6): p. 1523-7.

28. Walker-Smith, J.A., *Management of growth failure in Crohn's disease.* Arch Dis Child, 1996. **75**(4): p. 351-4.

29. Ballinger, A.B., et al., *Growth failure occurs through a decrease in insulin-like growth factor 1 which is independent of undernutrition in a rat model of colitis.* Gut, 2000. **46**(5): p. 694-700.

30. Louis, E., et al., *Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease.* Gut, 2001. **49**(6): p. 777-82.

31. Yang, H., K.D. Taylor, and J.I. Rotter, *Inflammatory bowel disease. I. Genetic epidemiology.* Mol Genet Metab, 2001. **74**(1-2): p. 1-21.

32. Gasche, C., et al., *A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998.* Inflamm Bowel Dis, 2000. **6**(1): p. 8-15.

33. Silverberg, M.S., et al., *Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology.* Can J Gastroenterol, 2005. **19 Suppl A**: p. 5-36.

34. Ahmad, T., et al., *The molecular classification of the clinical manifestations of Crohn's disease.* Gastroenterology, 2002. **122**(4): p. 854-66.

35. Sun, L., et al., *CARD15 genotype and phenotype analysis in 55 pediatric patients with Crohn disease from Saxony, Germany.* J Pediatr Gastroenterol Nutr, 2003. **37**(4): p. 492-7.

36. Orholm, M., et al., *Familial occurrence of inflammatory bowel disease.* N Engl J Med, 1991. **324**(2): p. 84-8.

37. Yang, H., et al., *Familial empirical risks for inflammatory bowel disease: differences between Jews and non-Jews.* Gut, 1993. **34**(4): p. 517-24.

38. Karban, A., et al., *Risk factors for perianal Crohn's disease: the role of genotype, phenotype, and ethnicity.* Am J Gastroenterol, 2007. **102**(8): p. 1702-8.

39. Kurata, J.H., et al., *Crohn's disease among ethnic groups in a large health maintenance organization.* Gastroenterology, 1992. **102**(6): p. 1940-8.

40. Blanchard, J.F., et al., *Small-area variations and sociodemographic correlates for the incidence of Crohn's disease and ulcerative colitis.* Am J Epidemiol, 2001. **154**(4): p. 328-35.

41. Montgomery, S.M., et al., *Asian ethnic origin and the risk of inflammatory bowel disease.* Eur J Gastroenterol Hepatol, 1999. **11**(5): p. 543-6.

42. Lee, Y.M., et al., *Racial differences in the prevalence of ulcerative colitis and Crohn's disease in Singapore.* J Gastroenterol Hepatol, 2000. **15**(6): p. 622-5.

43. Weterman, I.T. and A.S. Pena, *Familial incidence of Crohn's disease in The Netherlands and a review of the literature.* Gastroenterology, 1984. **86**(3): p. 449-52.

44. Tysk, C., et al., *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**(7): p. 990-6.

45. Halfvarson, J., et al., *Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics.* Gastroenterology, 2003. **124**(7): p. 1767-73.

46. Orholm, M., et al., *Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study.* Scand J Gastroenterol, 2000. **35**(10): p. 1075-81.

47. Thompson, N.P., et al., *Genetics versus environment in inflammatory bowel disease: results of a British twin study.* Bmj, 1996. **312**(7023): p. 95-6.

48. Montgomery, S.M., et al., *Prevalence of inflammatory bowel disease in British 26 year olds: national longitudinal birth cohort.* Bmj, 1998. **316**(7137): p. 1058-9.

49. Halme, L., et al., *Family and twin studies in inflammatory bowel disease.* World J Gastroenterol, 2006. **12**(23): p. 3668-72.

50. Simmons, R.A., *Developmental origins of diabetes: the role of epigenetic mechanisms.* Curr Opin Endocrinol Diabetes Obes, 2007. **14**(1): p. 13-6.

51. Henikoff, S. and M.A. Matzke, *Exploring and explaining epigenetic effects.* Trends Genet, 1997. **13**(8): p. 293-5.

52. Falls, J.G., et al., *Genomic imprinting: implications for human disease.* Am J Pathol, 1999. **154**(3): p. 635-47.

53. Jirtle, R.L., *Genomic imprinting and cancer.* Exp Cell Res, 1999. **248**(1): p. 18-24.

54. Beaudet, A.L., *Epigenetics and complex human disease: is there a role in IBD?* J Pediatr Gastroenterol Nutr, 2008. **46 Suppl 1**: p. E2.

55. Petronis, A. and R. Petroniene, *Epigenetics of inflammatory bowel disease.* Gut, 2000. **47**(2): p. 302-6.

56. Akolkar, P.N., et al., *Differences in risk of Crohn's disease in offspring of mothers and fathers with inflammatory bowel disease.* Am J Gastroenterol, 1997. **92**(12): p. 2241-4.

57. Price, W.H., *A high incidence of chronic inflammatory bowel disease in patients with Turner's syndrome.* J Med Genet, 1979. **16**(4): p. 263-6.

58. Hayward, P.A., J. Satsangi, and D.P. Jewell, *Inflammatory bowel disease and the X chromosome.* QJM, 1996. **89**(9): p. 713-8.

59. Hampe, J., et al., *Linkage of inflammatory bowel disease to human chromosome 6p.* Am J Hum Genet, 1999. **65**(6): p. 1647-55.

60. Mott, F.E., *Early Lung Cancer Action Project.* Lancet, 1999. **354**(9185): p. 1206-7.

61. Bayless, T.M., M.F. Picco, and M.C. LaBuda, *Genetic anticipation in Crohn's disease.* Am J Gastroenterol, 1998. **93**(12): p. 2322-5.

62. Colombel, J.F., D. Laharie, and B. Grandbastien, *Anticipating the onset of inflammatory bowel disease.* Gut, 1999. **44**(6): p. 773-4.

63. Ridley, R.M., et al., *Anticipation in Huntington's disease is inherited through the male line but may originate in the female.* J Med Genet, 1988. **25**(9): p. 589-95.

64. Heresbach, D., et al., *Anticipation in Crohn's disease may be influenced by gender and ethnicity of the transmitting parent.* Am J Gastroenterol, 1998. **93**(12): p. 2368-72.

65. Polito, J.M., 2nd, et al., *Preliminary evidence for genetic anticipation in Crohn's disease.* Lancet, 1996. **347**(9004): p. 798-800.

66. Lee, J.C., et al., *Why children with inflammatory bowel disease are diagnosed at a younger age than their affected parent.* Gut, 1999. **44**(6): p. 808-11.

67. Grandbastien, B., et al., *Anticipation in familial Crohn's disease.* Gut, 1998. **42**(2): p. 170-4.

68. Petronis, A., *The regulation of D2 dopamine receptor gene expression: epigenetic factors should not be forgotten.* Mol Psychiatry, 1999. **4**(3): p. 212-3.

69. Satsangi, J., D.P. Jewell, and J.I. Bell, *The genetics of inflammatory bowel disease.* Gut, 1997. **40**(5): p. 572-4.

70. Hugot, J.P., et al., *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease.* Nature, 2001. **411**(6837): p. 599-603.

71. Cavanaugh, J., *International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16.* Am J Hum Genet, 2001. **68**(5): p. 1165-71.

72. Satsangi, J., et al., *Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12.* Nat Genet, 1996. **14**(2): p. 199-202.

73. Cavanaugh, J.A., et al., *Analysis of Australian Crohn's disease pedigrees refines the localization for susceptibility to inflammatory bowel disease on chromosome 16.* Ann Hum Genet, 1998. **62**(Pt 4): p. 291-8.

74. Ogura, Y., et al., *Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB.* J Biol Chem, 2001. **276**(7): p. 4812-8.

75. Rioux, J.D., et al., *Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci.* Am J Hum Genet, 2000. **66**(6): p. 1863-70.

76. Ma, Y., et al., *A genome-wide search identifies potential new susceptibility loci for Crohn's disease.* Inflamm Bowel Dis, 1999. **5**(4): p. 271-8.

77. Duerr, R.H., et al., *High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12.* Am J Hum Genet, 2000. **66**(6): p. 1857-62.

78. Rioux, J.D., et al., *Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease.* Nat Genet, 2001. **29**(2): p. 223-8.

79. Cho, J.H., et al., *Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1.* Proc Natl Acad Sci U S A, 1998. **95**(13): p. 7502-7.

80. Hugot, J.P., et al., *Mapping of a susceptibility locus for Crohn's disease on chromosome 16.* Nature, 1996. **379**(6568): p. 821-3.

81. Hampe, J., et al., *Fine mapping of the chromosome 3p susceptibility locus in inflammatory bowel disease.* Gut, 2001. **48**(2): p. 191-7.

82. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease.* Nature, 2001. **411**(6837): p. 603-6.

83. Stokkers, P.C., et al., *HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis.* Gut, 1999. **45**(3): p. 395-401.

84. Brant, S.R. and Y.Y. Shugart, *Inflammatory bowel disease gene hunting by linkage analysis: rationale, methodology, and present status of the field.* Inflamm Bowel Dis, 2004. **10**(3): p. 300-11.

85. Hugot, J.P., H. Zouali, and S. Lesage, *Lessons to be learned from the NOD2 gene in Crohn's disease.* Eur J Gastroenterol Hepatol, 2003. **15**(6): p. 593-7.

86. Stoll, M., et al., *Genetic variation in DLG5 is associated with inflammatory bowel disease.* Nat Genet, 2004. **36**(5): p. 476-80.

87. Peltekova, V.D., et al., *Functional variants of OCTN cation transporter genes are associated with Crohn disease.* Nat Genet, 2004. **36**(5): p. 471-5.

88. Inohara, N., et al., *Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB.* J Biol Chem, 1999. **274**(21): p. 14560-7.

89. Daly, M.J., et al., *High-resolution haplotype structure in the human genome.* Nat Genet, 2001. **29**(2): p. 229-32.

90. Dring, M.M., et al., *The pregnane X receptor locus is associated with susceptibility to inflammatory bowel disease.* Gastroenterology, 2006. **130**(2): p. 341-8; quiz 592.

91. Potocnik, U., et al., *Polymorphisms in multidrug resistance 1 (MDR1) gene are associated with refractory Crohn disease and ulcerative colitis.* Genes Immun, 2004. **5**(7): p. 530-9.

92. Oostenbrug, L.E., et al., *Absence of association between the multidrug resistance (MDR1) gene and inflammatory bowel disease.* Scand J Gastroenterol, 2006. **41**(10): p. 1174-82.

93. Amre, D.K., et al., *Association between genetic variants in the IL-23R gene and early-onset Crohn's disease: results from a case-control and family-based study among Canadian children.* Am J Gastroenterol, 2008. **103**(3): p. 615-20.

94. Hirschhorn, J.N. and M.J. Daly, *Genome-wide association studies for common diseases and complex traits.* Nat Rev Genet, 2005. **6**(2): p. 95-108.

95. Duerr, R.H., et al., *A genome-wide association study identifies IL23R as an inflammatory bowel disease gene.* Science, 2006. **314**(5804): p. 1461-3.

96. Hampe, J., et al., *A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1.* Nat Genet, 2007. **39**(2): p. 207-11.

97. Libioulle, C., 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 Genet, 2007. **3**(4): p. e58.

98. Parkes, M., et al., *Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility.* Nat Genet, 2007. **39**(7): p. 830-2.

99. Rioux, J.D., et al., *Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis.* Nat Genet, 2007. **39**(5): p. 596-604.

100. Consortium, W.T.C.C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.

101. Calder, P.C., *Dietary arachidonic acid: harmful, harmless or helpful?* Br J Nutr, 2007. **98**(3): p. 451-3.

102. *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.

103. Pierik, M., et al., *The IBD international genetics consortium provides further evidence for linkage to IBD4 and shows gene-environment interaction.* Inflamm Bowel Dis, 2005. **11**(1): p. 1-7.

104. Bottini, N., et al., *A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes.* Nat Genet, 2004. **36**(4): p. 337-8.

105. Bottini, N., et al., *Type 2 diabetes and the genetics of signal transduction: a study of interaction between adenosine deaminase and acid phosphatase locus 1 polymorphisms.* Metabolism, 2004. **53**(8): p. 995-1001.

106. Zeggini, E. and M.I. McCarthy, *Identifying susceptibility variants for type 2 diabetes.* Methods Mol Biol, 2007. **376**: p. 235-50.

107. Criswell, L.A. and P.K. Gregersen, *Current understanding of the genetic aetiology of rheumatoid arthritis and likely future developments.* Rheumatology (Oxford), 2005. **44 Suppl 4**: p. iv9-iv13.

108. Kugathasan, S., et al., *Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease.* Nat Genet, 2008.

109. Iles, M.M., *What can genome-wide association studies tell us about the genetics of common disease?* PLoS Genet, 2008. **4**(2): p. e33.

110. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.

111. Freeman, H.J., *Comparison of longstanding pediatric-onset and adult-onset Crohn's disease.* J Pediatr Gastroenterol Nutr, 2004. **39**(2): p. 183-6.

112. Pilote, L., et al., *A comprehensive view of sex-specific issues related to cardiovascular disease.* CMAJ, 2007. **176**(6): p. S1-44.

113. Lahoz, C., et al., *Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study.* Atherosclerosis, 2001. **154**(3): p. 529-37.

114. Frikke-Schmidt, R., et al., *Gender- and age-specific contributions of additional DNA sequence variation in the 5' regulatory region of the APOE gene to prediction of measures of lipid metabolism.* Hum Genet, 2004. **115**(4): p. 331-45.

115. Ordovas, J.M., *Gender, a significant factor in the cross talk between genes, environment, and health.* Gend Med, 2007. **4 Suppl B**: p. S111-22.

116. Singh, P.P., M. Singh, and S.S. Mastana, *APOE distribution in world populations with new data from India and the UK.* Ann Hum Biol, 2006. **33**(3): p. 279-308.

117. Loughlin, J., et al., *Linkage analysis of chromosome 2q in osteoarthritis.* Rheumatology (Oxford), 2000. **39**(4): p. 377-81.

118. Fisher, S.A., et al., *Sex stratification of an inflammatory bowel disease genome search shows male-specific linkage to the HLA region of chromosome 6.* Eur J Hum Genet, 2002. **10**(4): p. 259-65.

119. Daly, M.J., et al., *Association of DLG5 R30Q variant with inflammatory bowel disease.* Eur J Hum Genet, 2005. **13**(7): p. 835-9.

120. Newman, W.G., et al., *DLG5 variants contribute to Crohn disease risk in a Canadian population.* Hum Mutat, 2006. **27**(4): p. 353-8.

121. Medici, V., et al., *Extreme heterogeneity in CARD15 and DLG5 Crohn disease-associated polymorphisms between German and Norwegian populations.* Eur J Hum Genet, 2006. **14**(4): p. 459-68.

122. Lakatos, P.L., et al., *DLG5 R30Q is not associated with IBD in Hungarian IBD patients but predicts clinical response to steroids in Crohn's disease.* Inflamm Bowel Dis, 2006. **12**(5): p. 362-8.

123. Ferraris, A., et al., *Relationship between CARD15, SLC22A4/5, and DLG5 polymorphisms and early-onset inflammatory bowel diseases: an Italian multicentric study.* Inflamm Bowel Dis, 2006. **12**(5): p. 355-61.

124. Buning, C., et al., *DLG5 variants in inflammatory bowel disease.* Am J Gastroenterol, 2006. **101**(4): p. 786-92.

125. Torok, H.P., et al., *Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease.* Gut, 2005. **54**(10): p. 1421-7.

126. Noble, C.L., et al., *DLG5 variants do not influence susceptibility to inflammatory bowel disease in the Scottish population.* Gut, 2005. **54**(10): p. 1416-20.

127. Friedrichs, F., et al., *Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men.* Hum Genet, 2006. **119**(3): p. 305-11.

128. Browning, B.L., et al., *Gender-stratified analysis of DLG5 R30Q in 4707 patients with Crohn disease and 4973 controls from 12 Caucasian cohorts.* J Med Genet, 2008. **45**(1): p. 36-42.

129. Biank, V., et al., *DLG5 R30Q variant is a female-specific protective factor in pediatric onset Crohn's disease.* Am J Gastroenterol, 2007. **102**(2): p. 391-8.

130. Farmer, R.G., W.A. Hawk, and R.B. Turnbull, Jr., *Clinical patterns in Crohn's disease: a statistical study of 615 cases.* Gastroenterology, 1975. **68**(4 Pt 1): p. 627-35.

131. Oostenbrug, L.E., et al., *Clinical outcome of Crohn's disease according to the Vienna classification: disease location is a useful predictor of disease course.* Eur J Gastroenterol Hepatol, 2006. **18**(3): p. 255-61.

132. Marion, J.F., *An expensive and empty ritual: the continued use of random biopsy technique for detecting dysplasia in patients with colitis--pro.* Inflamm Bowel Dis, 2007. **13**(10): p. 1271-2.

133. Sachar, D.B., et al., *Is perianal Crohn's disease associated with intestinal fistulization?* Am J Gastroenterol, 2005. **100**(7): p. 1547-9.

134. Annese, V., 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.* Am J Gastroenterol, 2001. **96**(10): p. 2939-45.

135. Lee, J.C. and J.E. Lennard-Jones, *Inflammatory bowel disease in 67 families each with three or more affected first-degree relatives.* Gastroenterology, 1996. **111**(3): p. 587-96.

136. Bayless, T.M., et al., *Crohn's disease: concordance for site and clinical type in affected family members--potential hereditary influences.* Gastroenterology, 1996. **111**(3): p. 573-9.

137. Louis, E., 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**(4): p. 552-7.

138. Brant, S.R., et al., *Defining complex contributions of NOD2/CARD15 gene mutations, age at onset, and tobacco use on Crohn's disease phenotypes.* Inflamm Bowel Dis, 2003. **9**(5): p. 281-9.

139. Vermeire, S., et al., *CARD15 genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure.* Am J Hum Genet, 2002. **71**(1): p. 74-83.

140. Lesage, S., et al., *CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease.* Am J Hum Genet, 2002. **70**(4): p. 845-57.

141. Hampe, J., et al., *Association of NOD2 (CARD 15) genotype with clinical course of Crohn's disease: a cohort study.* Lancet, 2002. **359**(9318): p. 1661-5.

142. Cuthbert, A.P., et al., *The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease.* Gastroenterology, 2002. **122**(4): p. 867-74.

143. Cummings, J.R. and D.P. Jewell, *Clinical implications of inflammatory bowel disease genetics on phenotype.* Inflamm Bowel Dis, 2005. **11**(1): p. 56-61.

144. Lala, S., et al., *Crohn's disease and the NOD2 gene: a role for paneth cells.* Gastroenterology, 2003. **125**(1): p. 47-57.

145. Economou, M., et al., *Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis.* Am J Gastroenterol, 2004. **99**(12): p. 2393-404.

146. Silverberg, M.S., et al., *A population- and family-based study of Canadian families reveals association of HLA DRB1*0103 with colonic involvement in inflammatory bowel disease.* Inflamm Bowel Dis, 2003. **9**(1): p. 1-9.

147. Armuzzi, A., et al., *Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31.* Gut, 2003. **52**(8): p. 1133-9.

148. de Ridder, L., et al., *Genetic susceptibility has a more important role in pediatric-onset Crohn's disease than in adult-onset Crohn's disease.* Inflamm Bowel Dis, 2007. **13**(9): p. 1083-92.

149. Nieuwenhuis, E.E. and J.C. Escher, *Early onset IBD: what's the difference?* Dig Liver Dis, 2008. **40**(1): p. 12-5.

150. Radford-Smith, G. and N. Pandeya, *Associations between NOD2/CARD15 genotype and phenotype in Crohn's disease--Are we there yet?* World J Gastroenterol, 2006. **12**(44): p. 7097-103.

151. Russell, R.K., 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**(8): p. 1114-23.

152. Silverberg, M.S., et al., *Diagnostic misclassification reduces the ability to detect linkage in inflammatory bowel disease genetic studies.* Gut, 2001. **49**(6): p. 773-6.

153. Moore, J.H., *The ubiquitous nature of epistasis in determining susceptibility to common human diseases.* Hum Hered, 2003. **56**(1-3): p. 73-82.

154. Moore, J.H. and S.M. Williams, *Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis.* Bioessays, 2005. **27**(6): p. 637-46.

155. Ritchie, M.D., *Bioinformatics approaches for detecting gene-gene and gene-environment interactions in studies of human disease.* Neurosurg Focus, 2005. **19**(4): p. E2.

156. Thornton-Wells, T.A., J.H. Moore, and J.L. Haines, *Genetics, statistics and human disease: analytical retooling for complexity.* Trends Genet, 2004. **20**(12): p. 640-7.

157. Okazaki, T., et al., *Contributions of IBD5, IL23R, ATG16L1, and NOD2 to Crohn's disease risk in a population-based case-control study: evidence of gene-gene interactions.* Inflamm Bowel Dis, 2008. **14**(11): p. 1528-41.

158. Peduzzi, P., et al., *A simulation study of the number of events per variable in logistic regression analysis.* J Clin Epidemiol, 1996. **49**(12): p. 1373-9.

159. Hosmer, D.W., et al., *A comparison of goodness-of-fit tests for the logistic regression model.* Stat Med, 1997. **16**(9): p. 965-80.

160. Moore, J.H. and S.M. Williams, *New strategies for identifying gene-gene interactions in hypertension.* Ann Med, 2002. **34**(2): p. 88-95.

161. Ritchie, M.D., et al., *Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer.* Am J Hum Genet, 2001. **69**(1): p. 138-47.

162. Ritchie, M.D., L.W. Hahn, and J.H. Moore, *Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity.* Genet Epidemiol, 2003. **24**(2): p. 150-7.

163. Silverberg, M.S., et al., *Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease.* Eur J Hum Genet, 2007. **15**(3): p. 328-35.

164. Gionchetti, P., et al., *Prophylaxis of pouchitis onset with probiotic therapy: a double-blind, placebo-controlled trial.* Gastroenterology, 2003. **124**(5): p. 1202-9.

165. Sutherland, L., et al., *Double blind, placebo controlled trial of metronidazole in Crohn's disease.* Gut, 1991. **32**(9): p. 1071-5.

166. Barnich, N., et al., *CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease.* J Clin Invest, 2007. **117**(6): p. 1566-74.

167. Darfeuille-Michaud, A., et al., *Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease.* Gastroenterology, 1998. **115**(6): p. 1405-13.

168. Onderdonk, A.B., J.A. Hermos, and J.G. Bartlett, *The role of the intestinal microflora in experimental colitis.* Am J Clin Nutr, 1977. **30**(11): p. 1819-25.

169. Elson, C.O., et al., *Experimental models of inflammatory bowel disease reveal innate, adaptive, and regulatory mechanisms of host dialogue with the microbiota.* Immunol Rev, 2005. **206**: p. 260-76.

170. Sellon, R.K., et al., *Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice.* Infect Immun, 1998. **66**(11): p. 5224-31.

171. Berg, R.D., *The indigenous gastrointestinal microflora.* Trends Microbiol, 1996. **4**(11): p. 430-5.

172. Gordon, J.I., et al., *Epithelial cell growth and differentiation. III. Promoting diversity in the intestine: conversations between the microflora, epithelium, and diffuse GALT.* Am J Physiol, 1997. **273**(3 Pt 1): p. G565-70.

173. Shanahan, F., *Probiotics and inflammatory bowel disease: is there a scientific rationale?* Inflamm Bowel Dis, 2000. **6**(2): p. 107-15.

174. Xavier, R.J. and D.K. Podolsky, *Unravelling the pathogenesis of inflammatory bowel disease.* Nature, 2007. **448**(7152): p. 427-34.

175. Buhner, S., et al., *Genetic basis for increased intestinal permeability in families with Crohn's disease: role of CARD15 3020insC mutation?* Gut, 2006. **55**(3): p. 342-7.

176. Irvine, E.J. and J.K. Marshall, *Increased intestinal permeability precedes the onset of Crohn's disease in a subject with familial risk.* Gastroenterology, 2000. **119**(6): p. 1740-4.

177. May, G.R., L.R. Sutherland, and J.B. Meddings, *Is small intestinal permeability really increased in relatives of patients with Crohn's disease?* Gastroenterology, 1993. **104**(6): p. 1627-32.

178. Soderholm, J.D., et al., *Augmented increase in tight junction permeability by luminal stimuli in the non-inflamed ileum of Crohn's disease.* Gut, 2002. **50**(3): p. 307-13.

179. Gassler, N., et al., *Inflammatory bowel disease is associated with changes of enterocytic junctions.* Am J Physiol Gastrointest Liver Physiol, 2001. **281**(1): p. G216-28.

180. Kabashima, K., et al., *The prostaglandin receptor EP4 suppresses colitis, mucosal damage and CD4 cell activation in the gut.* J Clin Invest, 2002. **109**(7): p. 883-93.

181. Kagnoff, M.F. and L. Eckmann, *Epithelial cells as sensors for microbial infection.* J Clin Invest, 1997. **100**(1): p. 6-10.

182. Hisamatsu, T., et al., *CARD15/NOD2 functions as an antibacterial factor in human intestinal epithelial cells.* Gastroenterology, 2003. **124**(4): p. 993-1000.

183. Di Carlo, E., et al., *The combined action of IL-15 and IL-12 gene transfer can induce tumor cell rejection without T and NK cell involvement.* J Immunol, 2000. **165**(6): p. 3111-8.

184. Yoshida, M., et al., *Neonatal Fc receptor for IgG regulates mucosal immune responses to luminal bacteria.* J Clin Invest, 2006. **116**(8): p. 2142-2151.

185. Neish, A.S., et al., *Prokaryotic regulation of epithelial responses by inhibition of IkappaB-alpha ubiquitination.* Science, 2000. **289**(5484): p. 1560-3.

186. Mashimo, H., et al., *Impaired defense of intestinal mucosa in mice lacking intestinal trefoil factor.* Science, 1996. **274**(5285): p. 262-5.

187. Wehkamp, J., et al., *NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression.* Gut, 2004. **53**(11): p. 1658-64.

188. Coombes, J.L. and K.J. Maloy, *Control of intestinal homeostasis by regulatory T cells and dendritic cells.* Semin Immunol, 2007. **19**(2): p. 116-26.

189. Zaph, C., et al., *Epithelial-cell-intrinsic IKK-beta expression regulates intestinal immune homeostasis.* Nature, 2007. **446**(7135): p. 552-6.

190. Becker, C., et al., *Constitutive p40 promoter activation and IL-23 production in the terminal ileum mediated by dendritic cells.* J Clin Invest, 2003. **112**(5): p. 693-706.

191. Hue, S., et al., *Interleukin-23 drives innate and T cell-mediated intestinal inflammation.* J Exp Med, 2006. **203**(11): p. 2473-83.

192. Mizoguchi, A., et al., *Dependence of intestinal granuloma formation on unique myeloid DC-like cells.* J Clin Invest, 2007. **117**(3): p. 605-15.

193. Smith, P.D., C. Ochsenbauer-Jambor, and L.E. Smythies, *Intestinal macrophages: unique effector cells of the innate immune system.* Immunol Rev, 2005. **206**: p. 149-59.

194. Takeda, K., et al., *Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of Stat3 in macrophages and neutrophils.* Immunity, 1999. **10**(1): p. 39-49.

195. Kamada, N., et al., *Abnormally differentiated subsets of intestinal macrophage play a key role in Th1-dominant chronic colitis through excess production of IL-12 and IL-23 in response to bacteria.* J Immunol, 2005. **175**(10): p. 6900-8.

196. Franchimont, D., et al., *Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis.* Gut, 2004. **53**(7): p. 987-92.

197. Goyette, P., et al., *Molecular pathogenesis of inflammatory bowel disease: genotypes, phenotypes and personalized medicine.* Ann Med, 2007. **39**(3): p. 177-99.

198. Pierik, M., et al., *Toll-like receptor-1, -2, and -6 polymorphisms influence disease extension in inflammatory bowel diseases.* Inflamm Bowel Dis, 2006. **12**(1): p. 1-8.

199. Torok, H.P., et al., *Crohn's disease is associated with a toll-like receptor-9 polymorphism.* Gastroenterology, 2004. **127**(1): p. 365-6.

200. Nieuwenhuis, E.E. and R.S. Blumberg, *The role of the epithelial barrier in inflammatory bowel disease.* Adv Exp Med Biol, 2006. **579**: p. 108-16.

201. Rescigno, M. and E.E. Nieuwenhuis, *The role of altered microbial signaling via mutant NODs in intestinal inflammation.* Curr Opin Gastroenterol, 2007. **23**(1): p. 21-6.

202. Chamaillard, M., et al., *Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation.* Cell Microbiol, 2003. **5**(9): p. 581-92.

203. McCole, D.F. and K.E. Barrett, *Varied role of the gut epithelium in mucosal homeostasis.* Curr Opin Gastroenterol, 2007. **23**(6): p. 647-54.

204. Wehkamp, J. and E.F. Stange, *A new look at Crohn's disease: breakdown of the mucosal antibacterial defense.* Ann N Y Acad Sci, 2006. **1072**: p. 321-31.

205. Gutierrez, M.G., et al., *Protective role of autophagy against Vibrio cholerae cytolysin, a pore-forming toxin from V. cholerae.* Proc Natl Acad Sci U S A, 2007. **104**(6): p. 1829-34.

206. Lee, H.K., et al., *Autophagy-dependent viral recognition by plasmacytoid dendritic cells.* Science, 2007. **315**(5817): p. 1398-401.

207. Schmid, D. and C. Munz, *Innate and adaptive immunity through autophagy.* Immunity, 2007. **27**(1): p. 11-21.

208. Xu, Y., et al., *Toll-like receptor 4 is a sensor for autophagy associated with innate immunity.* Immunity, 2007. **27**(1): p. 135-44.

209. Qu, X., et al., *Autophagy gene-dependent clearance of apoptotic cells during embryonic development.* Cell, 2007. **128**(5): p. 931-46.

210. 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**(7145): p. 661-78.

211. Singh, S.B., et al., *Human IRGM induces autophagy to eliminate intracellular mycobacteria.* Science, 2006. **313**(5792): p. 1438-41.

212. Taylor, G.A., *IRG proteins: key mediators of interferon-regulated host resistance to intracellular pathogens.* Cell Microbiol, 2007. **9**(5): p. 1099-107.

213. Mosmann, T.R., et al., *Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins.* J Immunol, 1986. **136**(7): p. 2348-57.

214. Murphy, T.J., et al., *CD4+CD25+ regulatory T cells control innate immune reactivity after injury.* J Immunol, 2005. **174**(5): p. 2957-63.

215. Weaver, C.T., et al., *IL-17 family cytokines and the expanding diversity of effector T cell lineages.* Annu Rev Immunol, 2007. **25**: p. 821-52.

216. Kastelein, R.A., C.A. Hunter, and D.J. Cua, *Discovery and biology of IL-23 and IL-27: related but functionally distinct regulators of inflammation.* Annu Rev Immunol, 2007. **25**: p. 221-42.

217. Bettelli, E., M. Oukka, and V.K. Kuchroo, *T(H)-17 cells in the circle of immunity and autoimmunity.* Nat Immunol, 2007. **8**(4): p. 345-50.

218. Kinugasa, T., et al., *Claudins regulate the intestinal barrier in response to immune mediators.* Gastroenterology, 2000. **118**(6): p. 1001-11.

219. Madsen, J.R., L.S. Laursen, and K. Lauritsen, *[Chronic inflammatory bowel disease].* Nord Med, 1992. **107**(10): p. 254-60.

220. Fritsche, K., *Fatty acids as modulators of the immune response.* Annu Rev Nutr, 2006. **26**: p. 45-73.

221. Funk, C.D., *Prostaglandins and leukotrienes: advances in eicosanoid biology.* Science, 2001. **294**(5548): p. 1871-5.

222. Gronert, K., et al., *Transcellular regulation of eicosanoid biosynthesis.* Methods Mol Biol, 1999. **120**: p. 119-44.

223. Simopoulos, A.P., *Omega-3 fatty acids in health and disease and in growth and development.* Am J Clin Nutr, 1991. **54**(3): p. 438-63.

224. Sharon, P. and W.F. Stenson, *Enhanced synthesis of leukotriene B4 by colonic mucosa in inflammatory bowel disease.* Gastroenterology, 1984. **86**(3): p. 453-60.

225. Lobos, E.A., P. Sharon, and W.F. Stenson, *Chemotactic activity in inflammatory bowel disease. Role of leukotriene B4.* Dig Dis Sci, 1987. **32**(12): p. 1380-8.

226. Lauritsen, K., et al., *In vivo profiles of eicosanoids in ulcerative colitis, Crohn's colitis, and Clostridium difficile colitis.* Gastroenterology, 1988. **95**(1): p. 11-7.

227. Schmidt, C., et al., *Arachidonic acid metabolism and intracellular calcium concentration in inflammatory bowel disease.* Eur J Gastroenterol Hepatol, 1995. **7**(9): p. 865-9.

228. Hendel, J., I. Ahnfelt-Ronne, and O.H. Nielsen, *Expression of 5-lipoxygenase mRNA is unchanged in the colon of patients with active inflammatory bowel disease.* Inflamm Res, 2002. **51**(8): p. 423-6.

229. Jupp, J., et al., *Colonic expression of leukotriene-pathway enzymes in inflammatory bowel diseases.* Inflamm Bowel Dis, 2007. **13**(5): p. 537-46.

230. Rask-Madsen, J., et al., *5-Lipoxygenase inhibitors for the treatment of inflammatory bowel disease.* Agents Actions, 1992. **Spec No**: p. C37-46.

231. Hawkey, C.J., et al., *A trial of zileuton versus mesalazine or placebo in the maintenance of remission of ulcerative colitis. The European Zileuton Study Group For Ulcerative Colitis.* Gastroenterology, 1997. **112**(3): p. 718-24.

232. Roberts, W.G., et al., *Leukotrienes in ulcerative colitis: results of a multicenter trial of a leukotriene biosynthesis inhibitor, MK-591.* Gastroenterology, 1997. **112**(3): p. 725-32.

233. De Caterina, R. and A. Zampolli, *From asthma to atherosclerosis--5-lipoxygenase, leukotrienes, and inflammation.* N Engl J Med, 2004. **350**(1): p. 4-7.

234. Kim, S.H., et al., *Leukotriene-related gene polymorphisms in patients with aspirin-intolerant urticaria and aspirin-intolerant asthma: differing contributions of ALOX5 polymorphism in Korean population.* J Korean Med Sci, 2005. **20**(6): p. 926-31.

235. Kim, S.H., et al., *Polymorphism of tandem repeat in promoter of 5-lipoxygenase in ASA-intolerant asthma: a positive association with airway hyperresponsiveness.* Allergy, 2005. **60**(6): p. 760-5.

236. Klotsman, M., et al., *Pharmacogenetics of the 5-lipoxygenase biosynthetic pathway and variable clinical response to montelukast.* Pharmacogenet Genomics, 2007. **17**(3): p. 189-96.

237. Taccone-Gallucci, M., et al., *N-3 PUFAs reduce oxidative stress in ESRD patients on maintenance HD by inhibiting 5-lipoxygenase activity.* Kidney Int, 2006. **69**(8): p. 1450-4.

238. Chen, X.S., et al., *Role of leukotrienes revealed by targeted disruption of the 5-lipoxygenase gene.* Nature, 1994. **372**(6502): p. 179-82.

239. Mehrabian, M., et al., *Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice.* Circ Res, 2002. **91**(2): p. 120-6.

240. Drazen, J.M., E. Israel, and P.M. O'Byrne, *Treatment of asthma with drugs modifying the leukotriene pathway.* N Engl J Med, 1999. **340**(3): p. 197-206.

241. Kalayci, O., et al., *ALOX5 promoter genotype, asthma severity and LTC production by eosinophils.* Allergy, 2006. **61**(1): p. 97-103.

242. Dwyer, J.H., et al., *Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis.* N Engl J Med, 2004. **350**(1): p. 29-37.

243. Drazen, J.M., et al., *Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment.* Nat Genet, 1999. **22**(2): p. 168-70.

244. Choi, J.H., et al., *Leukotriene-related gene polymorphisms in ASA-intolerant asthma: an association with a haplotype of 5-lipoxygenase.* Hum Genet, 2004. **114**(4): p. 337-44.

245. Kikuta, Y., E. Kusunose, and M. Kusunose, *Prostaglandin and leukotriene omega-hydroxylases.* Prostaglandins Other Lipid Mediat, 2002. **68-69**: p. 345-62.

246. Murphy, R.C. and M.A. Gijon, *Biosynthesis and metabolism of leukotrienes.* Biochem J, 2007. **405**(3): p. 379-95.

247. Kikuta, Y., E. Kusunose, and M. Kusunose, *Characterization of human liver leukotriene B(4) omega-hydroxylase P450 (CYP4F2).* J Biochem, 2000. **127**(6): p. 1047-52.

248. van Heel, D.A., et al., *Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs.* Hum Mol Genet, 2004. **13**(7): p. 763-70.

249. Low, J.H., et al., *Inflammatory bowel disease is linked to 19p13 and associated with ICAM-1.* Inflamm Bowel Dis, 2004. **10**(3): p. 173-81.

250. Babron, M.C., et al., *Meta and pooled analysis of European coeliac disease data.* Eur J Hum Genet, 2003. **11**(11): p. 828-34.

251. Van Belzen, M.J., et al., *A major non-HLA locus in celiac disease maps to chromosome 19.* Gastroenterology, 2003. **125**(4): p. 1032-41.

252. Curley, C.R., et al., *A functional candidate screen for coeliac disease genes.* Eur J Hum Genet, 2006. **14**(11): p. 1215-22.

253. Tello-Ruiz, M.K., et al., *Haplotype-based association analysis of 56 functional candidate genes in the IBD6 locus on chromosome 19.* Eur J Hum Genet, 2006. **14**(6): p. 780-90.

254. Barrett, J.C., et al., *Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease.* Nat Genet, 2008. **40**(8): p. 955-62.

255. Curtis, D., *Use of siblings as controls in case-control association studies.* Ann Hum Genet, 1997. **61 (Pt 4)**: p. 319-33.

256. Spielman, R.S., R.E. McGinnis, and W.J. Ewens, *Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM).* Am J Hum Genet, 1993. **52**(3): p. 506-16.

257. Falk, C.T. and P. Rubinstein, *Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations.* Ann Hum Genet, 1987. **51**(Pt 3): p. 227-33.

258. Spielman, R.S. and W.J. Ewens, *A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test.* Am J Hum Genet, 1998. **62**(2): p. 450-8.

259. Schaid, D.J. and S.S. Sommer, *Genotype relative risks: methods for design and analysis of candidate-gene association studies.* Am J Hum Genet, 1993. **53**(5): p. 1114-26.

260. Nagelkerke, N.J., et al., *Combining the transmission disequilibrium test and case-control methodology using generalized logistic regression.* Eur J Hum Genet, 2004. **12**(11): p. 964-70.

261. Epstein, M.P., et al., *Genetic association analysis using data from triads and unrelated subjects.* Am J Hum Genet, 2005. **76**(4): p. 592-608.

262. Schaid, D.J., *Case-parents design for gene-environment interaction.* Genet Epidemiol, 1999. **16**(3): p. 261-73.

263. Niu, T., *Algorithms for inferring haplotypes.* Genet Epidemiol, 2004. **27**(4): p. 334-47.

264. Umbach, D.M. and C.R. Weinberg, *Designing and analysing case-control studies to exploit independence of genotype and exposure.* Stat Med, 1997. **16**(15): p. 1731-43.

265. Umbach, D.M. and C.R. Weinberg, *The use of case-parent triads to study joint effects of genotype and exposure.* Am J Hum Genet, 2000. **66**(1): p. 251-61.

266. Umbach, D.M., *Invited commentary: on studying the joint effects of candidate genes and exposures.* Am J Epidemiol, 2000. **152**(8): p. 701-3.

267. Lennard-Jones, J.E., *Classification of inflammatory bowel disease.* Scand J Gastroenterol Suppl, 1989. **170**: p. 2-6; discussion 16-9.

268. Sands, B.E., *From symptom to diagnosis: clinical distinctions among various forms of intestinal inflammation.* Gastroenterology, 2004. **126**(6): p. 1518-32.

269. Aps, J.K., et al., *Flow cytometry as a new method to quantify the cellular content of human saliva and its relation to gingivitis.* Clin Chim Acta, 2002. **321**(1-2): p. 35-41.

270. Greinix, H.T., B. Volc-Platzer, and R. Knobler, *Criteria for assessing chronic GVHD.* Bone Marrow Transplant, 2000. **25**(5): p. 575.

271. Steinberg, K., et al., *DNA banking for epidemiologic studies: a review of current practices.* Epidemiology, 2002. **13**(3): p. 246-54.

272. Ross, P., et al., *High level multiplex genotyping by MALDI-TOF mass spectrometry.* Nat Biotechnol, 1998. **16**(13): p. 1347-51.

273. Carlson, C.S., et al., *Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium.* Am J Hum Genet, 2004. **74**(1): p. 106-20.

274. Meylan, E., J. Tschopp, and M. Karin, *Intracellular pattern recognition receptors in the host response.* Nature, 2006. **442**(7098): p. 39-44.

275. Schurmann, M., et al., *CARD15 gene mutations in sarcoidosis.* Eur Respir J, 2003. **22**(5): p. 748-54.

276. Steer, S., et al., *Development of rheumatoid arthritis is not associated with two polymorphisms in the Crohn's disease gene CARD15.* Rheumatology (Oxford), 2003. **42**(2): p. 304-7.

277. Douglas, J.A., A.D. Skol, and M. Boehnke, *Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data.* Am J Hum Genet, 2002. **70**(2): p. 487-95.

278. Pompanon, F., et al., *Genotyping errors: causes, consequences and solutions.* Nat Rev Genet, 2005. **6**(11): p. 847-59.

279. Schaid, D.J. and S.S. Sommer, *Comparison of statistics for candidate-gene association studies using cases and parents.* Am J Hum Genet, 1994. **55**(2): p. 402-9.

280. Purcell, S., M.J. Daly, and P.C. Sham, *WHAP: haplotype-based association analysis.* Bioinformatics, 2007. **23**(2): p. 255-6.

281. Schaid, D.J., et al., *Caution on pedigree haplotype inference with software that assumes linkage equilibrium.* Am J Hum Genet, 2002. **71**(4): p. 992-5.

282. Akaike, H., *[Data analysis by statistical models].* No To Hattatsu, 1992. **24**(2): p. 127-33.

283. Moore, J.H., *Computational analysis of gene-gene interactions using multifactor dimensionality reduction.* Expert Rev Mol Diagn, 2004. **4**(6): p. 795-803.

284. Motsinger, A.A. and M.D. Ritchie, *Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies.* Hum Genomics, 2006. **2**(5): p. 318-28.

285. Hahn, L.W., M.D. Ritchie, and J.H. Moore, *Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions.* Bioinformatics, 2003. **19**(3): p. 376-82.

286. Motsinger-Reif, A.A., et al., *A comparison of analytical methods for genetic association studies.* Genet Epidemiol, 2008. **32**(8): p. 767-78.

287. Devlin, B. and N. Risch, *A comparison of linkage disequilibrium measures for fine-scale mapping.* Genomics, 1995. **29**(2): p. 311-22.

288. Skol, A.D., et al., *Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies.* Nat Genet, 2006. **38**(2): p. 209-13.

289. Epstein, M.P., I.D. Waldman, and G.A. Satten, *Improved association analyses of disease subtypes in case-parent triads.* Genet Epidemiol, 2006. **30**(3): p. 209-19.

290. Crooks, S.W. and R.A. Stockley, *Leukotriene B4.* Int J Biochem Cell Biol, 1998. **30**(2): p. 173-8.

291. Cuzzocrea, S., et al., *5-Lipoxygenase modulates colitis through the regulation of adhesion molecule expression and neutrophil migration.* Lab Invest, 2005. **85**(6): p. 808-22.

292. Rubin, P. and K.W. Mollison, *Pharmacotherapy of diseases mediated by 5-lipoxygenase pathway eicosanoids.* Prostaglandins Other Lipid Mediat, 2007. **83**(3): p. 188-97.

293. In, K.H., et al., *Naturally occurring mutations in the human 5-lipoxygenase gene promoter that modify transcription factor binding and reporter gene transcription.* J Clin Invest, 1997. **99**(5): p. 1130-7.

294. Matter, K. and M.S. Balda, *Functional analysis of tight junctions.* Methods, 2003. **30**(3): p. 228-34.

295. Marks, D.J., et al., *Defective acute inflammation in Crohn's disease: a clinical investigation.* Lancet, 2006. **367**(9511): p. 668-78.

296. Dieckgraefe, B.K., et al., *Association of glycogen storage disease 1b and Crohn disease: results of a North American survey.* Eur J Pediatr, 2002. **161 Suppl 1**: p. S88-92.

297. Ishii, E., et al., *Chediak-Higashi syndrome with intestinal complication. Report of a case.* J Clin Gastroenterol, 1987. **9**(5): p. 556-8.

298. D'Agata, I.D., et al., *Leucocyte adhesion deficiency presenting as a chronic ileocolitis.* Gut, 1996. **39**(4): p. 605-8.

299. Korzenik, J.R., et al., *Sargramostim for active Crohn's disease.* N Engl J Med, 2005. **352**(21): p. 2193-201.

300. Freeman, H.J., *Application of the Vienna Classification for Crohn's disease to a single clinician database of 877 patients.* Can J Gastroenterol, 2001. **15**(2): p. 89-93.

301. Cosnes, J., et al., *Long-term evolution of disease behavior of Crohn's disease.* Inflamm Bowel Dis, 2002. **8**(4): p. 244-50.

302. Ogura, Y., et al., *Expression of NOD2 in Paneth cells: a possible link to Crohn's ileitis.* Gut, 2003. **52**(11): p. 1591-7.

303. Sullivan, P.F., *Spurious genetic associations.* Biol Psychiatry, 2007. **61**(10): p. 1121-6.

304. Smith, G.D. and S. Ebrahim, *Mendelian randomization: prospects, potentials, and limitations.* Int J Epidemiol, 2004. **33**(1): p. 30-42.

305. Schwartz, J.H., *Recognizing William Bateson's contributions.* Science, 2007. **315**(5815): p. 1077.

APPENDIX 1

Crohn’s disease (CD) patients and controls recruitment and acquisition of clinical information

Phase 1 - Screening Study (case-parent design)	**Phase 2 - Replication Study (case-control design)**	
	Cases	Controls
Pediatric case-population from: Hopital Ste. Justine (HSJ), British Columbia Children's Hospital (BCCH), and The Children's Hospital of Eastern Ontario (CHEO)	Pediatric case-population from: Hopital Ste. Justine (HSJ), and The Children's Hospital of Eastern Ontario (CHEO)	Source population: 1. Pediatric Orthopedic Departments – HSJ and CHEO; 2. Healthy adults from Saguenay-Lac-Saint-Jean region; 3. Newborns.
A. Identification of patients diagnosed with CD at the 3 hospitals: 1. HSJ (2003-2007): incident + prevalent cases; 2. BCCH (2004): incident cases; 3. CHEO (2006-2007): incident cases.	A. Identification of patients diagnosed with CD at the 3 hospitals: 1. HSJ (2003-2007): incident + prevalent cases; 2. CHEO (2006-2007): incident cases.	Selection of controls: 1. Children with minor traumas; 2. Participants in a population genetics study at HSJ; 3. Cord Blood Biobank at HSJ.
B. Confirmation of diagnosis (minimal follow-up 12 months or latest diagnosis for new patients)	B. Confirmation of diagnosis (minimal follow-up 12 months or latest diagnosis for new patients)	Exclusions: 1. Children with immune mediated diseases (asthma, juvenile diabetes, celiac disease etc); 2. None; 3. None.
C. Acquired clinical information from medical charts and biological samples were collected.	C. Acquired clinical information from medical charts	Cascade recruitment of controls: 1. Siblings of healthy children with minor traumas.
D. Patients with Ulcerative Colitis (UC) and Indeterminate Colitis (IC) excluded	D. Patients with Ulcerative Colitis (UC) and Indeterminate Colitis (IC) excluded	Information collection
E. Patients and parents contacted and invited for participation	E. Patients contacted and invited for participation	
F. Information collection	F. Information collection	

APPENDIX 2

Relevant clinical information acquired from Crohn's disease (CD) patients

CLINICAL INFORMATION FOR CROHN'S DISEASE (CD) PATIENTS

Date completed: ______/_______/_______
Day Month Year

Study ID: ______________
Age: _______/_______
Years Months
Sex: ____________

Diagnosis:
Ulcerative Colitis :______ Crohn's Disease______ Indeterminate colitis_____
Date of Diagnosis: ____/_____/_____
Day month year

CLINICAL DATA

Intestinal manifestations

	Yes	**No**	**Unknown**	
Urgency		____	____	____
Nocturnal diarrhea	____	____	____	
Incontinence	____	____	____	
Tenesmus	____	____	____	
Diarrhoea	____	____	____	
Abdominal pain		____	____	____
Rectal bleeding		____	____	____
Fistula:				
Perianal		____	____	____
Enteroenteric	____	____	____	
Enterocutaneous		____	____	____
Rectovaginal	____	____	____	
Enterovesicular		____	____	____
Enterolabial	____	____	____	
Other: _____________	____	____	____	
Fissure	____	____	____	
Internal abscess		____	____	____
Perianal abscess		____	____	____
Others ___________________	____	____	____	
___________________	____	____	____	

Extraintestinal manifestations

Growth parameters Weight ____________% Height ________________%

	Yes	No	Unknown	
Skin				
Erythema Nodosum	____	____	____	
Pyoderma Gangrenosum		____	____	____
Psoriasis		____	____	____

Eyes

Iritis	____	____	____
Conjuctivitis	____	____	____
Uveitis	____	____	____
Episcleritis	____	____	____

Musculoskeletal

Arthritis		____	____	____
Sacroileitis	____	____	____	
Arthralgia	____	____	____	
Ankylosing Spondylitis		____	____	____
Clubbing	____	____	____	

Liver

Viral hepatitis	____	____	____	
Sclerosing Cholangitis		____	____	____
Autoimmune hepatitis	____	____	____	
Nonspecific elevated liver tests		____	____	____
Other ________________________		____	____	____

Miscellaneous

Gallstones	____	____	____	
Aphthae		____	____	____
Nephrolithiasis		____	____	____
Pouchitis	____	____	____	
Malnutrition	____	____	____	
Obstruction; site _____________		____	____	____
Osteoporosis/Osteomalacia	____	____	____	
Fatigue	____	____	____	

Disease location (Anatomic extent of disease)

Oropharyngeal ___ Esophagus ____ Stomach ____
Duodenum ____

Jejunum ____ Ileum ____ Cecum ____ Ascending colon ____

Transverse colon ____ Descending colon ____ Sigmoid colon ____ Rectum ____

Perianal ____ Anal ____

Disease behavior

Stricturing ____ (colon ____, small bowel ____) UC-like ____
Internal perforating ____ (colon ____, small bowel ____) Perianal perforating ____
Non perforating/non stricturing ____

Disease activity

	Yes	No	
No activity	____	____	
Mild activity	____	____	
Moderate activity		____	____
High activity	____	____	
Fulminate activity			

Laboratory data

CBC with manual differentiation: WBC__________ Hgb________ MCV _______ PLT _______

ABS Polys _____________ ABS lymphs __________

CRP ______________ ESR _______________ Albumin ___________

Vit. B12 _____________ Vitamin D _____________

RBC folate _____________ Amylase ___________ Lipase ___________

ALT _____________ AST _______________ Total bilirubin ____________

Alkaline phosphatase _________ Iron __________

DIAGNOSTIC DATA

Endoscopic studies

				Disease Location													
Procedure	Date	BX	Hosp	OP	E	S	D	J	I	C	AC	TV	DC	SC	R	P	A

(BX: Barium X-ray; OP: oropharynx; E: eosophagus; S: Stomach; D: Duodenum; J: Jejunum; I: ileum; C: caecum; AC: ascending colon; TV: transverse colon; D: descending colon; SC: sigmoid colon; R: rectum; P: perinanal; A: anal)

Histopathologic studies

	Disease Location													
Findings	OP	E	S	D	J	I	C	AC	TV	DC	SC	R	P	A
CD compatible														
Granulomas														
UC compatible														
Indeterminate colitis														

(OP: oropharynx; E: eosophagus; S: Stomach; D: Duodenum; J: Jejunum; I: ileum; C: caecum; AC: ascending colon; TV: transverse colon; D: descending colon; SC: sigmoid colon; R: rectum; P: perinanal; A: anal)

Radiologic studies

				Disease Location													
Procedure	Date	Hosp.	Finding	OP	E	S	D	J	I	C	AC	TV	DC	SC	R	P	A

(H: Hospital; OP: oropharynx; E: eosophagus; S: Stomach; D: Duodenum; J: Jejunum; I: ileum; C: caecum; AC: ascending colon; TV: transverse colon; D: descending colon; SC: sigmoid colon; R: rectum; P: perinanal; A: anal)

MANAGEMENT DATA

Surgery procedures

# of IBD surgeries ________

Small bowel resection with ileostomy ____ without ileostomy ____

Large bowel resection segmental with colostomy ____ segmental without colostomy ____

subtotal colectomy with colostomy ____ subtotal colectomy without colostomy ____

total colectomy with colostomy ____ total colectomy with ileoanal pull-through ____

Stricturoplasties ____
Asbcess drainage ____
Indication Bleeding ____ Perforation ____ Obstruction ____ Other ____
(____________________)

Hospitalization

Indication ____
Intervention ____

Initial medical management/Drug-related adverse events

Allergies **No ____** **Yes ____** **Specify**
____

1. Mesalamine

Asacol Previous ____ , date started ____________ date stopped _________ dose ________
Current ____ , date started ____________ dose ________
Side effects No ____ Yes ____ Specify:
____

Pentasa Previous ____ , date started ____________ date stopped _________ dose ________
Current ____ , date started ____________ dose ________
Side effects No ____ Yes ____ Specify:
____

Sulfasalazine Previous ____ , date started ____________ date stopped _________ dose ________
Current ____ , date started ____________ dose ________
Side effects No ____ Yes ____ Specify:
____

Balsalazide Previous ____ , date started ____________ date stopped _________ dose ________
Current ____ , date started ____________ dose ________
Side effects No ____ Yes ____ Specify:
____

Rowasa enema Previous ____ , date started ____________ date stopped _________ dose ________
Current ____ , date started ____________ dose ________

Side effects No ____ Yes ____ Specify:
____

Rowasa suppository
Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose ________
Side effects No ____ Yes ____ Specify:
____

Corticosteroids **Never used steroids ____**

Prednisone First course ____ , date started ____________ date stopped _________
dose ________
Disease activity when steroids started: Mild Moderate Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Second course ____ , date started ____________ dose
____
Disease activity when steroids started: Mild Moderate Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Third course ____ , date started ____________ dose
________ Disease activity when steroids started: Mild Moderate
Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Budesonide First course ____ , date started ____________ date stopped _________
dose _________
Disease activity when steroids started: Mild Moderate Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Second course ____ , date started ____________ dose
____
Disease activity when steroids started: Mild Moderate Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Third course ____ , date started ____________ dose
________ Disease activity when steroids started: Mild Moderate
Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Other steroid medication
First course ____ , date started ____________ date stopped _________ dose _________
Disease activity when steroids started: Mild Moderate Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Second course ____ , date started ____________ dose
____
Disease activity when steroids started: Mild Moderate Severe
Responded to steroids: Yes No

Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Third course ____ , date started ____________ dose
________ Disease activity when steroids started: Mild Moderate
Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Cortisone enema/foam
First course ____ , date started ____________ date stopped _________
dose ________
Disease activity when steroids started: Mild Moderate Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Second course ____ , date started ____________ dose
____
Disease activity when steroids started: Mild Moderate Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No
If yes, date of surgery:

Third course ____ , date started ____________ dose
________ Disease activity when steroids started: Mild Moderate
Severe
Responded to steroids: Yes No
Able to taper off steroids Yes ____ No ____
Side effects No ____ Yes ____ Specify:
____
Underwent surgery Yes No

If yes, date of surgery:

Immunomodulator therapy

Imuran Previous ____ , date started ____________ date stopped _________ dose
____
Current ____ , date started ____________ dose
____
Side effects No ____ Yes ____ Specify:
____

6 MP Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose
____
Side effects No ____ Yes ____ Specify:
____
TPMT testing performed Yes ____ No ____

Cyclosporine Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose
____
Side effects No ____ Yes ____ Specify:
____

MTX Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose
____
Side effects No ____ Yes ____ Specify:
____

FK 506 Previous ____ , date started ____________ date stopped
_________ dose ________
Current ____ , date started ____________ dose
____
Side effects No ____ Yes ____ Specify:
____

Celcept/MMF Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose
____

Side effects No ____ Yes ____ Specify:
____

Infliximab Usage Yes____ No____
Date of infusion _______________ Improvement No____ Yes____
Date of infusion _______________ Improvement No____ Yes____
Date of infusion _______________ Improvement No____ Yes____
Date of infusion _______________ Improvement No____ Yes____
Date of infusion _______________ Improvement No____ Yes____

Antibiotics

Metronidazole Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose
____
Side effects No____ Yes____ Specify:
____

Ciprofloxacin Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose
____
Side effects No____ Yes____ Specify:
____

Neomycin Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose
____
Side effects No____ Yes____ Specify:
____

Pain medication

Name: ________________________________
Previous ____ , date started ____________ date stopped _________
dose ________
Current ____ , date started ____________ dose
____
Side effects No____ Yes____ Specify:
____

Anti-diarrheal medication

Name: ________________________________

Previous ____ , date started ____________ date stopped _________ dose ________

Current ____ , date started ____________ dose ________

Side effects No____ Yes____ Specify: ____________________

TPN Yes____ No____

Enteral

Nutrition Yes____ No____ Name ________________________

Supplements Name (vitamins, herbs, probiotics) ______________________________

Previous ____ , date started ____________ date stopped _________ dose ________

Current ____ , date started ____________ dose ________

Side effects No____ Yes____ Specify: ____________________

APPENDIX 3

Tag – SNP selection sequence for CYP4F2 and ALOX5 genes

Gene Name: **CYP4F2**

Gene ID: 8529

Chromosome 19: 15849833 - 15869884 (-)

Allele Frequency Cutoff (%): 0

R2 Threshold for Clusters: 0.8

Minimal Genotype Coverage (%) of Snps to Be TagSnps: 85

Minimal Genotype Coverage (%) of Snps to Be Clustered: 70

Data Merging: common samples with combined variations

Population: PGA-EUROPEAN-PANEL, Submitter: PGA-UW-FHCRC

Bin	Total Number of Sites	Average Minor Allele Frequency	Tag SNPs	Other SNPs
1	25	11 %	rs2074902 rs3093098 rs3093173	rs2006193 rs2365178 rs3093100 rs3093103 rs3093105 rs3093106 rs3093110 rs3093112 rs3093114 rs3093115 rs3093116 rs3093120 rs3093121 rs3093122 rs3093124 rs3093128 rs3093129 rs3093134 rs3093135 rs3093160 rs3093167 rs3093180
2	6	47 %	rs1126433 rs2074900 rs3093097 rs3093194	rs3093195

			rs3093203	
3	6	39 %	rs3093145 rs3093148 rs3093159 rs3093170 rs3093182	rs3093156
4	4	24 %	rs1272 rs758154 rs3093184 rs3093190	
5	4	5 %	rs2215092 rs3093150 rs3093153 rs3093224	
6	3	27 %	rs3093158 rs3093166 rs3093168	
7	2	21 %	rs3093193 rs3093199	
8	2	7 %	rs3093200 rs3093204	
9	2	4 %	rs3093220 rs3093222	
10	1	0 %	rs736089	
11	1	11 %	rs2016503	
12	1	17 %	rs2108622	
13	1	7 %	rs2886296	
14	1	0 %	rs3093093	
15	1	0 %	rs3093094	
16	1	0 %	rs3093095	
17	1	0 %	rs3093096	
18	1	0 %	rs3093099	
19	1	0 %	rs3093101	
20	1	0 %	rs3093102	
21	1	0 %	rs3093104	
22	1	0 %	rs3093107	
23	1	0 %	rs3093108	
24	1	0 %	rs3093109	
25	1	0 %	rs3093111	
26	1	0 %	rs3093113	
27	1	0 %	rs3093117	
28	1	0 %	rs3093118	
29	1	0 %	rs3093119	

30	1	0 %	rs3093123	
31	1	0 %	rs3093125	
32	1	0 %	rs3093126	
33	1	0 %	rs3093127	
34	1	0 %	rs3093130	
35	1	0 %	rs3093131	
36	1	0 %	rs3093132	
37	1	0 %	rs3093133	
38	1	0 %	rs3093136	
39	1	0 %	rs3093137	
40	1	0 %	rs3093138	
41	1	0 %	rs3093139	
42	1	0 %	rs3093140	
43	1	0 %	rs3093141	
44	1	0 %	rs3093143	
45	1	10 %	rs3093144	
46	1	0 %	rs3093146	
47	1	0 %	rs3093147	
48	1	0 %	rs3093149	
49	1	0 %	rs3093151	
50	1	0 %	rs3093152	
51	1	0 %	rs3093154	
52	1	0 %	rs3093155	
53	1	0 %	rs3093157	
54	1	0 %	rs3093161	
55	1	0 %	rs3093162	
56	1	0 %	rs3093163	
57	1	0 %	rs3093165	
58	1	0 %	rs3093171	
59	1	0 %	rs3093172	
60	1	0 %	rs3093174	
61	1	0 %	rs3093175	
62	1	0 %	rs3093176	
63	1	0 %	rs3093177	
64	1	0 %	rs3093178	
65	1	0 %	rs3093179	
66	1	0 %	rs3093181	
67	1	0 %	rs3093183	
68	1	0 %	rs3093185	
69	1	0 %	rs3093186	
70	1	0 %	rs3093188	
71	1	0 %	rs3093189	
72	1	0 %	rs3093191	

73	1	0 %	rs3093192	
74	1	0 %	rs3093196	
75	1	0 %	rs3093197	
76	1	30 %	rs3093198	
77	1	0 %	rs3093201	
78	1	2 %	rs3093218	
79	1	2 %	rs3093219	
80	1	4 %	rs3093223	
81	1	2 %	rs3093225	
82	1	7 %	rs3093226	
83	1	2 %	rs3093227	
84	1	0 %	rs3093142	
85	1	3 %	rs3093221	

Variation Color code:
splice-site
coding-nonsynonymous
coding-synonymous
coding
mrna-utr

Gene Name: **ALOX5**

Gene ID: 240

Chromosome 10: 45189634 - 45261567 (+)

Allele Frequency Cutoff (%): 0

R2 Threshold for Clusters: 0.8

Minimal Genotype Coverage (%) of Snps to Be TagSnps: 85

Minimal Genotype Coverage (%) of Snps to Be Clustered: 70

Data Merging: common samples with combined variations

Population: AFD_EUR_PANEL, Submitter: PERLEGEN

Bin	Total Number of Sites	Average Minor Allele Frequency	Tag SNPs	Other SNPs
1	5	13 %	rs3824613 rs12783095 rs17157728 rs17157733 rs17157736	
2	3	15 %	rs745986 rs11239501 rs17444064	
3	3	2 %	rs7077173 rs7917687 rs17157771	
4	2	43 %	rs1369214 rs2115819	
5	2	12 %	rs1487562 rs7080474	
6	2	32 %	rs1565096 rs17523178	
7	2	43 %	rs3780901 rs7090328	
8	1	2 %	rs2228064	
9	1	6 %	rs2288619	
10	1	42 %	rs2291427	
11	1	2 %	rs3780905	
12	1	38 %	rs7080713	
13	1	0 %	rs10128306	

14	1	48 %	rs10751383	
15	1	0 %	rs11239499	
16	1	0 %	rs11239500	
17	1	0 %	rs11239502	
18	1	0 %	rs11239503	
19	1	0 %	rs11239504	
20	1	38 %	rs12264801	
21	1	0 %	rs17153289	
22	1	0 %	rs17157731	
23	1	0 %	rs17157756	
24	1	0 %	rs17157784	
25	1	4 %	rs17522720	

Variation Color code
splice-site
coding-nonsynonymous
coding-synonymous
coding
mrna-utr

APPENDIX 4 A+B

A - Models of power calculations for the case-parent design

B - Models of power calculations for the case-control only design

Summary Table:

Allele frequency	**Alpha**	**Power**	**Model**	**Relative risk**	**Trios required**
10%	0.05	80%	Additive	2.0	141
			Additive	1.8	203
			Recessive	2.0	1587
			Recessive	4.0	271
			Dominant	2.0	175
			Dominant	2.2	133
20%	0.05	80%	Additive	2.0	87
			Additive	1.7	152
			Recessive	2.0	448
			Recessive	3.0	147
			Dominant	2.0	130
			Dominant	1.8	165
30%	0.05	80%	Additive	2.0	71
			Additive	1.5	213
			Additive	1.7	123
			Recessive	2.0	233
			Recessive	2.5	123
			Dominant	2.0	130
			Dominant	1.8	177
50%	0.05	80%	Additive	2.0	69
			Additive	1.6	146
			Recessive	2.0	129
			Recessive	1.8	182
			Dominant	2.0	189
			Dominant	2.2	152

```
Model # 1
Outcome:                    Disease
Design:                   Case-parent
Hypothesis:                 Gene only
Desired power:                0.800000
Significance level:           0.050000, 2-sided
Gene

 Mode of inheritance:        Log-additive
   Allele frequency:          0.1000
Disease model                 Summary parameters
    P0:     0.000100            *kp:     0.000121
    RG:       2.0000            (*indicates calculated value)

Parameter                   N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                    141    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 2
Outcome:                    Disease
Design:                   Case-parent
Hypothesis:                 Gene only
Desired power:                0.800000
Significance level:           0.050000, 2-sided
Gene
   Mode of inheritance:        Log-additive
   Allele frequency:          0.1000
Disease model                 Summary parameters
    P0:     0.000100            *kp:     0.000117
   RG:      1.8000              (*indicates calculated value)

Parameter                N    Null      Full      Reduced
-----------------------------------------------------------------
Gene                   203    bG=0      bG        ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 3
Outcome:                      Disease
Design:                     Case-parent
Hypothesis:                   Gene only
Desired power:                  0.800000
Significance level:             0.050000, 2-sided
Gene
  Mode of inheritance:          Dominant
  Allele frequency:             0.1000
Disease model                   Summary parameters
   P0:    0.000100               *kp:     0.000119
   RG:      2.0000              (*indicates calculated value)

Parameter                N    Null      Full      Reduced
-----------------------------------------------------------------
Gene                   175    bG=0      bG        ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 4
Outcome:                      Disease
Design:                     Case-parent
Hypothesis:                   Gene only
```

```
Desired power:              0.800000
Significance level:         0.050000, 2-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:         0.1000
Disease model               Summary parameters
   P0:    0.000100            *kp:    0.000123
   RG:      2.2000            (*indicates calculated value)

Parameter                 N    Null      Full      Reduced
-------------------------------------------------------------
Gene                    133   bG=0      bG        ----
-------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 5
Outcome:                     Disease
Design:                     Case-parent
Hypothesis:                  Gene only
Desired power:              0.800000
Significance level:         0.050000, 2-sided
Gene
  Mode of inheritance:        Recessive
  Allele frequency:         0.1000
Disease model               Summary parameters
   P0:    0.000100            *kp:    0.000101
   RG:      2.0000            (*indicates calculated value)

Parameter                 N    Null      Full      Reduced
-------------------------------------------------------------
Gene                   1587   bG=0      bG        ----
```

```
----------------------------------------------------------------
N is the number of case-parent trios required for the desired power


Model # 6
Outcome:                    Disease
Design:                   Case-parent
Hypothesis:                 Gene only
Desired power:                0.800000
Significance level:           0.050000, 2-sided
Gene
  Mode of inheritance:        Recessive
  Allele frequency:           0.1000
Disease model                 Summary parameters
   P0:    0.000100             *kp:    0.000103
   RG:      4.0000             (*indicates calculated value)

Parameter                 N    Null      Full      Reduced
----------------------------------------------------------------
Gene                    271    bG=0      bG        ----
----------------------------------------------------------------
N is the number of case-parent trios required for the desired power



Model # 7
Outcome:                    Disease
Design:                   Case-parent
Hypothesis:                 Gene only
Desired power:                0.800000
Significance level:           0.050000, 2-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:          0.2000
Disease model                Summary parameters
   P0:    0.000100             *kp:    0.000144
   RG:      2.0000             (*indicates calculated value)

Parameter                N    Null      Full      Reduced
--------------------------------------------------------------
Gene                    87    bG=0      bG        ----
--------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 8
Outcome:                       Disease
Design:                      Case-parent
Hypothesis:                    Gene only
Desired power:                   0.800000
Significance level:             0.050000, 2-sided
Gene
  Mode of inheritance:         Log-additive
  Allele frequency:          0.2000
Disease model                Summary parameters
   P0:    0.000100             *kp:    0.000130
   RG:      1.7000             (*indicates calculated value)

Parameter                N    Null      Full      Reduced
--------------------------------------------------------------
Gene                   152    bG=0      bG        ----
--------------------------------------------------------------
N is the number of case-parent trios required for the desired power
```

```
Model # 9
Outcome:                  Disease
Design:                  Case-parent
Hypothesis:               Gene only
Desired power:              0.800000
Significance level:         0.050000, 2-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:         0.2000
Disease model               Summary parameters
   P0:    0.000100            *kp:    0.000136
   RG:      2.0000            (*indicates calculated value)

Parameter               N    Null      Full      Reduced
---------------------------------------------------------------
Gene                  130    bG=0      bG         ----
---------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 10
Outcome:                  Disease
Design:                  Case-parent
Hypothesis:               Gene only
Desired power:              0.800000
Significance level:         0.050000, 2-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:         0.2000
Disease model               Summary parameters
   P0:    0.000100            *kp:    0.000125
```

```
    RG:      1.7000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                     222    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 11
Outcome:                        Disease
Design:                       Case-parent
Hypothesis:                     Gene only
Desired power:                    0.800000
Significance level:               0.050000, 2-sided
Gene
   Mode of inheritance:         Dominant
   Allele frequency:              0.2000
Disease model                     Summary parameters
    P0:     0.000100               *kp:     0.000131
    RG:      1.8500              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                     165    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 12
Outcome:                        Disease
Design:                       Case-parent
Hypothesis:                     Gene only
```

```
Desired power:              0.800000
Significance level:         0.050000, 2-sided
Gene
  Mode of inheritance:        Recessive
  Allele frequency:         0.2000
Disease model               Summary parameters
   P0:    0.000100            *kp:    0.000104
   RG:      2.0000            (*indicates calculated value)

Parameter              N   Null     Full     Reduced
----------------------------------------------------------------
Gene                 448   bG=0     bG        ----
----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 13
Outcome:                    Disease
Design:                    Case-parent
Hypothesis:                 Gene only
Desired power:              0.800000
Significance level:         0.050000, 2-sided
Gene
  Mode of inheritance:        Recessive
  Allele frequency:         0.2000
Disease model               Summary parameters
   P0:    0.000100            *kp:    0.000108
   RG:      3.0000            (*indicates calculated value)

Parameter              N   Null     Full     Reduced
----------------------------------------------------------------
Gene                 147   bG=0     bG        ----
```

```
------------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 14
Outcome:                    Disease
Design:                   Case-parent
Hypothesis:                 Gene only
Desired power:                0.800000
Significance level:           0.050000, 2-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:           0.3000
Disease model                 Summary parameters
   P0:    0.000100              *kp:    0.000169
   RG:      2.0000              (*indicates calculated value)

Parameter                 N    Null      Full      Reduced
------------------------------------------------------------------
Gene                    71    bG=0      bG        ----
------------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 15
Outcome:                    Disease
Design:                   Case-parent
Hypothesis:                 Gene only
Desired power:                0.800000
Significance level:           0.050000, 2-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:           0.3000
```

```
Disease model              Summary parameters
   P0:    0.000100           *kp:    0.000132
   RG:      1.5000           (*indicates calculated value)

Parameter                N    Null      Full      Reduced
-------------------------------------------------------------
Gene                   213    bG=0      bG        ----
-------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 16
Outcome:                     Disease
Design:                    Case-parent
Hypothesis:                  Gene only
Desired power:                 0.800000
Significance level:            0.050000, 2-sided
Gene
  Mode of inheritance:         Log-additive
  Allele frequency:            0.3000
Disease model              Summary parameters
   P0:    0.000100           *kp:    0.000146
   RG:      1.7000           (*indicates calculated value)

Parameter                N    Null      Full      Reduced
-------------------------------------------------------------
Gene                   123    bG=0      bG        ----
-------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 17
Outcome:                     Disease
```

```
Design:                    Case-parent
Hypothesis:                  Gene only
Desired power:                 0.800000
Significance level:            0.050000, 2-sided
Gene
  Mode of inheritance:          Dominant
  Allele frequency:            0.3000
Disease model                  Summary parameters
   P0:     0.000100              *kp:     0.000151
   RG:       2.0000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                     130    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 18
Outcome:                     Disease
Design:                    Case-parent
Hypothesis:                  Gene only
Desired power:                 0.800000
Significance level:            0.050000, 2-sided
Gene
  Mode of inheritance:          Dominant
  Allele frequency:            0.3000
Disease model                  Summary parameters
   P0:     0.000100              *kp:     0.000141
   RG:       1.8000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
```

```
----------------------------------------------------------------
Gene                    177    bG=0       bG         ----
----------------------------------------------------------------
N is the number of case-parent trios required for the desired power


Model # 19
Outcome:                        Disease
Design:                       Case-parent
Hypothesis:                     Gene only
Desired power:                    0.800000
Significance level:               0.050000, 2-sided
Gene
  Mode of inheritance:          Recessive
  Allele frequency:               0.3000
Disease model                     Summary parameters
   P0:     0.000100                 *kp:     0.000109
   RG:       2.0000                  (*indicates calculated value)


Parameter                     N    Null       Full       Reduced
----------------------------------------------------------------
Gene                    233    bG=0       bG         ----
----------------------------------------------------------------
N is the number of case-parent trios required for the desired power


Model # 20
Outcome:                        Disease
Design:                       Case-parent
Hypothesis:                     Gene only
Desired power:                    0.800000
Significance level:               0.050000, 2-sided
Gene
```

```
  Mode of inheritance:        Recessive
  Allele frequency:          0.3000
Disease model                Summary parameters
   P0:     0.000100            *kp:     0.000118
   RG:       3.0000            (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                    80    bG=0       bG        ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 21
Outcome:                      Disease
Design:                     Case-parent
Hypothesis:                   Gene only
Desired power:                  0.800000
Significance level:             0.050000, 2-sided
Gene
  Mode of inheritance:        Recessive
  Allele frequency:          0.3000
Disease model                Summary parameters
   P0:     0.000100            *kp:     0.000114
   RG:       2.5000            (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                   123    bG=0       bG        ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power
```

```
Model # 22
Outcome:                    Disease
Design:                    Case-parent
Hypothesis:                 Gene only
Desired power:               0.800000
Significance level:          0.050000, 2-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:          0.5000
Disease model                Summary parameters
   P0:    0.000100             *kp:    0.000225
   RG:      2.0000             (*indicates calculated value)

Parameter                N    Null      Full      Reduced
-----------------------------------------------------------------
Gene                   69    bG=0      bG        ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 17
Outcome:                    Disease
Design:                    Case-parent
Hypothesis:                 Gene only
Desired power:               0.800000
Significance level:          0.050000, 2-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:          0.5000
Disease model                Summary parameters
   P0:    0.000100             *kp:    0.000156
```

```
    RG:      1.5000              (*indicates calculated value)

Parameter                 N    Null      Full      Reduced
--------------------------------------------------------------
Gene                    195    bG=0      bG         ----
--------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 18
Outcome:                      Disease
Design:                      Case-parent
Hypothesis:                   Gene only
Desired power:                  0.800000
Significance level:             0.050000, 2-sided
Gene
  Mode of inheritance:          Log-additive
  Allele frequency:             0.5000
Disease model                   Summary parameters
    P0:    0.000100              *kp:    0.000169
    RG:      1.6000              (*indicates calculated value)

Parameter                 N    Null      Full      Reduced
--------------------------------------------------------------
Gene                    146    bG=0      bG         ----
--------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 19
Outcome:                      Disease
Design:                      Case-parent
Hypothesis:                   Gene only
```

```
Desired power:              0.800000
Significance level:          0.050000, 2-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:          0.5000
Disease model               Summary parameters
   P0:     0.000100           *kp:     0.000175
   RG:       2.0000            (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                    189    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of case-parent trios required for the desired power

Model # 20
Outcome:                     Disease
Design:                    Case-parent
Hypothesis:                  Gene only
Desired power:              0.800000
Significance level:          0.050000, 2-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:          0.5000
Disease model               Summary parameters
   P0:     0.000100           *kp:     0.000190
   RG:       2.2000            (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                    152    bG=0       bG         ----
```

```
---------------------------------------------------------------
N is the number of case-parent trios required for the desired power


Model # 21
Outcome:                    Disease
Design:                    Case-parent
Hypothesis:                  Gene only
Desired power:                0.800000
Significance level:            0.050000, 2-sided
Gene
  Mode of inheritance:         Recessive
  Allele frequency:            0.5000
Disease model                 Summary parameters
   P0:     0.000100              *kp:     0.000125
   RG:       2.0000              (*indicates calculated value)


Parameter                  N    Null       Full       Reduced
---------------------------------------------------------------
Gene                     129    bG=0       bG         ----
---------------------------------------------------------------
N is the number of case-parent trios required for the desired power


Model # 22
Outcome:                    Disease
Design:                    Case-parent
Hypothesis:                  Gene only
Desired power:                0.800000
Significance level:            0.050000, 2-sided
Gene
  Mode of inheritance:         Recessive
  Allele frequency:            0.5000
```

```
Disease model              Summary parameters
   P0:    0.000100           *kp:    0.000120
   RG:      1.8000           (*indicates calculated value)

Parameter                  N    Null      Full      Reduced
------------------------------------------------------------
Gene                     182    bG=0      bG        ----
------------------------------------------------------------
N is the number of case-parent trios required for the desired power
```

Summary Table:

Allele frequency	Alpha (1-sided)	Power	Model	Relative risk	Cases*
22%	0.05	80%	Additive	1.6	121
			Additive	1.4	242
			Dominant	2.0	175
			Dominant	2.2	133
35%	0.05	80%	Additive	1.3	318
			Additive	1.4	192
			Dominant	1.6	203
			Dominant	1.5	270
45%	0.05	80%	Additive	1.3	301
			Additive	1.5	127
			Dominant	2.0	130
			Dominant	1.8	177

*Based on a case-control ratio of ~1:1.5. The allele frequencies correspond to the following CYP4F2 SNPs: rs1272 (22%), rs3093158 (35%) and rs3093145 (45%), that were estimated in the pseudo-controls in the screening case-parent study.

Dominant Model:

Model # 1

Outcome:	Disease
Design:	Unmatched case-control (1:1.5)
Hypothesis:	Gene only
Desired power:	0.800000
Significance level:	0.050000, 1-sided
Gene	
Mode of inheritance:	Dominant
Allele frequency:	0.1000
Disease model	Summary parameters

```
   P0:     0.000100             *kp:     0.000119
   RG:       2.0000             (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                     115    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 2
Outcome:                      Disease
Design:                     Unmatched case-control (1:1.5)
Hypothesis:                   Gene only
Desired power:                  0.800000
Significance level:             0.050000, 1-sided
Gene
  Mode of inheritance:          Dominant
  Allele frequency:             0.1000
Disease model                   Summary parameters
   P0:     0.000100             *kp:     0.000109
   RG:       1.5000             (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                     360    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN
```

```
Model # 3
Outcome:                    Disease
Design:                     Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:              0.800000
Significance level:         0.050000, 1-sided
Gene
  Mode of inheritance:      Dominant
  Allele frequency:         0.1000
Disease model               Summary parameters
   P0:    0.000100            *kp:    0.000113
   RG:      1.7000            (*indicates calculated value)

Parameter               N    Null     Full     Reduced
-------------------------------------------------------------------
Gene                  203    bG=0     bG       ----
-------------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 4
Outcome:                    Disease
Design:                     Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:              0.800000
Significance level:         0.050000, 1-sided
Gene
  Mode of inheritance:      Dominant
  Allele frequency:         0.2200
Disease model               Summary parameters
   P0:    0.000100            *kp:    0.000139
```

```
    RG:      2.0000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
----------------------------------------------------------------
Gene                      87    bG=0       bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 5
Outcome:                        Disease
Design:                       Unmatched case-control (1:1.5)
Hypothesis:                     Gene only
Desired power:                    0.800000
Significance level:               0.050000, 1-sided
Gene
   Mode of inheritance:          Dominant
   Allele frequency:              0.2200
Disease model                    Summary parameters
    P0:     0.000100              *kp:     0.000120
    RG:      1.5000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
----------------------------------------------------------------
Gene                     255    bG=0       bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN
```

```
Model # 6
Outcome:                    Disease
Design:                    Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:               0.800000
Significance level:           0.050000, 1-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:           0.2200
Disease model                Summary parameters
   P0:    0.000100            *kp:    0.000123
   RG:      1.6000            (*indicates calculated value)

Parameter                N    Null      Full      Reduced
----------------------------------------------------------------
Gene                   190   bG=0       bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 7
Outcome:                    Disease
Design:                    Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:               0.800000
Significance level:           0.050000, 1-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:           0.4500
Disease model                Summary parameters
   P0:    0.000100            *kp:    0.000135
```

```
    RG:      1.5000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                    328    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 8
Outcome:                        Disease
Design:                       Unmatched case-control (1:1.5)
Hypothesis:                     Gene only
Desired power:                    0.800000
Significance level:               0.050000, 1-sided
Gene
   Mode of inheritance:         Dominant
   Allele frequency:            0.4500
Disease model                   Summary parameters
    P0:     0.000100              *kp:     0.000156
    RG:      1.8000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                    164    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 9
Outcome:                        Disease
```

```
Design:                     Unmatched case-control (1:1.5)
Hypothesis:                   Gene only
Desired power:                  0.800000
Significance level:             0.050000, 1-sided
Gene
  Mode of inheritance:           Dominant
  Allele frequency:             0.4500
Disease model                   Summary parameters
   P0:     0.000100               *kp:     0.000142
   RG:       1.6000               (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                     248    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN
```

Multiplicative Model:

```
Model # 1
Outcome:                      Disease
Design:                     Unmatched case-control (1:1.5)
Hypothesis:                   Gene only
Desired power:                  0.800000
Significance level:             0.050000, 1-sided
Gene
  Mode of inheritance:           Log-additive
  Allele frequency:             0.2200
Disease model                   Summary parameters
   P0:     0.000100               *kp:     0.000138
   RG:      1.8000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
----------------------------------------------------------------
Gene                      75    bG=0       bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 2
Outcome:                      Disease
Design:                     Unmatched case-control (1:1.5)
Hypothesis:                   Gene only
Desired power:                  0.800000
Significance level:             0.050000, 1-sided
Gene
   Mode of inheritance:         Log-additive
   Allele frequency:            0.2200
Disease model                   Summary parameters
    P0:     0.000100              *kp:     0.000128
    RG:      1.6000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
----------------------------------------------------------------
Gene                     121    bG=0       bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN
```

```
Model # 3
Outcome:                    Disease
Design:                     Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:              0.800000
Significance level:         0.050000, 1-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:         0.2200
Disease model               Summary parameters
   P0:     0.000100           *kp:     0.000118
   RG:      1.4000            (*indicates calculated value)

Parameter                N    Null      Full      Reduced
-----------------------------------------------------------------
Gene                  242   bG=0      bG        ----
-----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 4
Outcome:                    Disease
Design:                     Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:              0.800000
Significance level:         0.050000, 1-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:         0.3500
Disease model               Summary parameters
   P0:     0.000100           *kp:     0.000130
```

```
    RG:      1.4000              (*indicates calculated value)

Parameter                   N    Null       Full       Reduced
----------------------------------------------------------------
Gene                      192    bG=0       bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 5
Outcome:                        Disease
Design:                       Unmatched case-control (1:1.5)
Hypothesis:                     Gene only
Desired power:                    0.800000
Significance level:               0.050000, 1-sided
Gene
   Mode of inheritance:          Log-additive
   Allele frequency:              0.3500
Disease model                     Summary parameters
    P0:     0.000100               *kp:     0.000122
    RG:      1.3000              (*indicates calculated value)

Parameter                   N    Null       Full       Reduced
----------------------------------------------------------------
Gene                      318    bG=0       bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN
```

```
Model # 6
Outcome:                    Disease
Design:                   Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:                0.800000
Significance level:           0.050000, 1-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:           0.3500
Disease model                 Summary parameters
   P0:     0.000100             *kp:     0.000146
   RG:       1.8000              (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
----------------------------------------------------------------
Gene                    133    bG=0       bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 7
Outcome:                    Disease
Design:                   Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:                0.800000
Significance level:           0.050000, 1-sided
Gene
  Mode of inheritance:        Dominant
  Allele frequency:           0.3500
Disease model                 Summary parameters
   P0:     0.000100             *kp:     0.000129
```

```
    RG:      1.5000              (*indicates calculated value)

Parameter                  N    Null      Full      Reduced
----------------------------------------------------------------
Gene                     270    bG=0      bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 8
Outcome:                       Disease
Design:                      Unmatched case-control (1:1.5)
Hypothesis:                    Gene only
Desired power:                   0.800000
Significance level:              0.050000, 1-sided
Gene
   Mode of inheritance:          Dominant
   Allele frequency:             0.3500
Disease model                    Summary parameters
    P0:     0.000100               *kp:     0.000135
    RG:      1.6000              (*indicates calculated value)

Parameter                  N    Null      Full      Reduced
----------------------------------------------------------------
Gene                     203    bG=0      bG         ----
----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN
```

```
Model # 9
Outcome:                    Disease
Design:                    Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:               0.800000
Significance level:          0.050000, 1-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:          0.4500
Disease model                Summary parameters
   P0:     0.000100            *kp:     0.000150
   RG:       1.5000             (*indicates calculated value)

Parameter                  N    Null       Full       Reduced
-----------------------------------------------------------------
Gene                   127    bG=0       bG         ----
-----------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN

Model # 10
Outcome:                    Disease
Design:                    Unmatched case-control (1:1.5)
Hypothesis:                 Gene only
Desired power:               0.800000
Significance level:          0.050000, 1-sided
Gene
  Mode of inheritance:        Log-additive
  Allele frequency:          0.4500
Disease model                Summary parameters
   P0:     0.000100            *kp:     0.000129
   RG:      1.3000            (*indicates calculated value)

Parameter                N    Null      Full      Reduced
----------------------------------------------------------------------
Gene                   301   bG=0       bG         ----
----------------------------------------------------------------------
N is the number of cases required for the desired power
The required number of controls is 1.5xN
```