# Genetic investigation of inflammatory bowel disease and post-infectious irritable bowel syndrome: the contribution of innate immunity candidate risk variants

Alexandra-Chloé Villani

Faculty of Medicine,

The Division of Experimental Medicine

McGill University, Montréal

February, 2009

A thesis submitted to McGill University

in partial fulfilment of the requirements of the degree of

Doctor of Philiosophy

*Your file Votre référence*
*ISBN:* 978-0-494-66673-9
*Our file Notre référence*
*ISBN:* 978-0-494-66673-9

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF SUPPLEMENTARY FIGURES



# LIST OF SUPPLEMENTARY TABLES

# ABSTRACT

The gastro-intestinal (GI) tract represents the largest surface of the body and is continuously exposed to the microbial environment. In such anatomy, the survival of the host requires that the intestinal microbial flora be contained without excessive immune-reactivity to commensal bacteria while retaining the ability to respond to episodic pathogens. The discriminative recognition between beneficial commensal bacteria and potentially harmful pathogens demands an accurate interpretation by the GI mucosal immune system. Any defects in the processes of innate immune recognition and killing may lead to the development and perpetuation of chronic intestinal inflammation, namely inflammatory bowel disease (i.e. Crohn's disease (CD) and ulcerative colitis (UC)) and post-infectious irritable bowel syndrome (PI-IBS). The aim of ours studies was to evaluate the contribution of candidate genes, involved in the homeostasis and regulation of the intestinal innate immune response, to the susceptibility to CD, UC, and PI-IBS. In the first phase, we describe functional and genetic association results supporting *NLRP3*, encoding NALP3/cryopyrin, as a novel CD susceptibility gene. We subsequently report that the *MEFV* gene, encoding pyrin, known to interact with and be involved in the same pathway as NALP3/cryopyrin, does not contribute to CD and UC susceptibility. No CD or UC additional associations were observed upon *NLRP3-MEFV* gene-gene interaction analyses. In the third phase, we report the first association study evaluating genetic determinants for PI-IBS, using the well-characterized Walkerton population cohort. We uncovered variants in the *TLR9*, *CDH1*, and *IL6* regions associated with PI-IBS

susceptibility. These results are in keeping with the pathophysiologic changes observed in patients with PI-IBS, which include increased intestinal permeability and intestinal immune activation.

Overall, these results contribute to a better understanding of the genetic susceptibility to CD, UC and PI-IBS and shed light on new pathogenic signaling pathways in the development of these diseases.

# RÉSUMÉ

Le tube digestif, représentant la plus grande surface du corps humain, est constamment exposé à un environnement microbactérien. Dans une telle anatomie, la survie de l'hôte requiert que la flore microbacterienne soit préservée sans réaction immunitaire excessive en la présence de cette flore tout en retenant sa capacité à répondre efficacement à la présence épisodique de pathogènes. La reconnaissance discriminatoire entre la flore bactérienne bénéfique et les pathogènes potentiellement nocifs requiert une interprétation précise de la part du système immunitaire digestif. Tous mécanismes défectueux dans la reconnaissance et destruction peuvent mener au développement et à la perpétuation de maladies inflammatoires chroniques du système digestif, notamment aux maladies inflammatoires de l'intestin (i.e. la malade de Crohn (MC) et la colite ulcéreuse (CU)) et au syndrome de l'intestin irritable post-infectieux (SII-PI). L'objectif premier des études présentées dans cette thèse était d'évaluer la contribution de gènes candidats impliqués dans l'homéostasie et la régulation de la réponse immunitaire innée de l'intestin à la susceptibilité de la MC, la CU, et le SII-PI. Dans la première phase, nous décrivons des résultats d'études d'association génétiques et fonctionnelles supportant *NLRP3*, encodant NALP3/cryopyrin, comme nouveau gène de susceptibilité de la MC. Nous démontrons ensuite que le gène *MEFV*, qui encode pour la protéine pyrin, qui interagi et est impliqué dans les mêmes cascades de signalement que NALP3/ cryopyrin, ne contribue pas à la susceptibilité de la MC et la CU. De plus, aucunes associations supplémentaires avec la MC ou la CU ne furent observées

suite à des analyses d'interactions gène-gène entre *NLRP3-MEFV*. Enfin, nous rapportons la première étude d'association évaluant le rôle de déterminants génétiques dans le SII-PI en utilisant la cohorte de population provenant de Walkerton. Nous décrivons des associations entre les régions *TLR9*, *CDH1*, et *IL6* et la susceptibilité au SII-PI. Ces résultats sont consistants avec les changements pathophysiologiques observés chez ces patients, incluant une augmentation de la perméabilité intestinale et l'activation immunitaire dans la muqueuse du colon.

En résumé, nos résultats contribuent à améliorer les connaissances sur la susceptibilité à la MC, la CU, et le SII-PI, et suggèrent de nouvelles cascades de signalement potentiellement impliquées dans le développement de ces maladies.

# ACKNOWLEDGEMENTS

I would like to first acknowledge my supervisor, Dr. Denis Franchimont, for taking me into his lab, sharing with me his passion for research, and for all his support and guidance throughout my studies despite our long distance relationship. I would also like to thank Dr. Thomas J. Hudson for welcoming me at the Genome Centre, for your time, teaching and mentoring, and for opening my eyes to the world of genetics! A special thank to Mathieu Lemire, my mentor, teacher, and friend. You were my pillar when I thought everything would fall apart when Drs. Franchimont and Hudson left Montreal; I could not have done it without you. Thanks for sticking with me from Toronto! Thank you to Dr. Marshall for trusting me with leading the Walkerton project. Thank you Genny, my dear fellow PhD colleague, for our numerous insightful discussions and your support on the journey of being a student supervised in a long-distance manner. To my second family at the Genome Centre, thank you for having adopted me and having supported me throughout my ups and downs; I was never really left alone thanks to you all! To Alexandre Montpetit and Tony Kwan, thank you for having tirelessly answered all my questions, cheered me up, and corrected all my papers throughout the years! There are many research colleagues who deserve mention. I would especially like to thank Genny Fortin, Catherine Collette, Alexandre Belisle, Yannick Renaud, and Sébastien Brunet for their insight and technical help. Thank you to Kris, Sandra, Annie-Claude, Lisa-Marie, Stéphanie, Alex, Amy, Steve, Ariane, Philippe, Jade, and Micheline for your support and friendship throughout these years. Merci à ma famille: Zara, papa et maman pour

votre support moral pendant toutes ces années, pour avoir toujours eu confiance en moi et de ne pas avoir douté de mes capacités à réussir. Un merci tout spécial à Jean-François pour ton support inconditionnel dans mes choix et décisions, et pour avoir accepté de vivre à distance pendant ces 4 dernières années en attendant que je termine mes études. These projects could not have been possible without the collaboration of many clinicians and collaborators who collected countless samples from diagnosed patients. My sincerest appreciations as well to the patients who have agreed to participate to these studies. Lastly, I would like to thank the Quebec government who provided me with a FRSQ Ph.D. scholarship.

# CONTRIBUTION OF AUTHORS

## Chapter 2

Villani AC, Lemire M, Fortin G, Louis E, Silverberg MS, Collette C, Baba N, Libioulle C, Belaiche J, Bitton A, Gaudet D, Cohen A, Langelier D, Fortin PR, Wither JE, Sarfati M, Rutgeerts P, Rioux JD, Vermeire S, Hudson TJ, Franchimont D. Common variants in the NLRP3 region contribute to Crohn disease susceptibility. *Nat Genet* 2009; 41: 71-76.

- Villani AC: Conceived, designed, and performed all the genotyping, sequencing, and expression experiments, including RNA and DNA extraction of monocytes. Performed all statistical analyses and wrote the manuscript.
- Lemire M: Supervised and corrected statistical analyses, prepared the figures, and revised the manuscript.
- Fortin G: Took care of colitis mice study, isolated monocytes from healthy donors, and revised the manuscript.
- Baba N: Isolated monocytes from healthy donors.
- Provided studied samples: Louis E, Silverberg MS, Collette C, Libioulle Belaiche J, Bitton A, Gaudet D, Cohen A, Langelier D, Fortin PR, Wither JE, Sarfati M, Rutgeerts P, Rioux JD, Vermeire S.
- Hudson TJ and Franchimont D: Study conception, supervision and manuscript revision.

## Chapter 3

Villani AC, Lemire M, Louis E, Silverberg MS, Collette C, Nimmo ER, Renaud Y, Brunet S, Fortin G, Belaiche J, Bitton A, Gaudet D, Cohen A, Rioux JD, Arnott IDR, Rutgeerts P, Satsangi J, Vermeire S, Hudson TJ, Franchimont D. Genetic variation in the familial Mediterranean fever gene (*MEFV*) and risk for inflammatory bowel disease. *PLoS ONE* (submitted).

- Villani AC: Conceived, designed, and performed all the genotyping, sequencing experiments. Performed statistical analyses and wrote the manuscript.
- Lemire M: Supervised, performed, and corrected statistical analyses, prepared the figures for association analyses, and revised the manuscript.
- Collette C: Performed expression experiments and help with genotyping experiments.
- Fortin G: Took care of colitis mice study.
- Provided studied samples: Louis E, Silverberg MS, Nimmo ER, Renaud Y, Brunet S, Belaiche J, Bitton A, Gaudet D, Cohen A, Rioux JD, Arnott IDR, Rutgeerts P, Satsangi J, Vermeire S.
- Hudson TJ and Franchimont D: Study conception, supervision and manuscript revision.

## Chapter 4

Title: *NLRP3-MEFV* gene-gene interactions do not contribute to inflammatory bowel disease susceptibility (*unpublished data*).

- Villani AC: Performed the genotyping experiments and some of the statistical analyses, and wrote the section.
- Lemire M: Performed and corrected statistical analyses, prepared the figures, and revised the section.

## Chapter 5

Villani AC, Lemire M, Thabane M, Geneau G, Belisle A, Garg AX, Clark WF, Moayyedi P, Collins SM, Franchimont D, Marshall JK. Genetic risk factors for post-infectious irritable bowel syndrome following a waterborne outbreak of gastroenteritis. *JAMA* (submitted).

- Villani AC: Managed DNA isolation of the cohort, performed literature review, selected candidate genes, conceived and designed all the experiments, helped in the genotyping experiments, performed some of the statistical analyses and wrote the manuscript.
- Lemire M: Supervised, performed, and corrected statistical analyses, prepared the figures for association analyses, and revised the manuscript.
- Thabane M: Managed the epidemiological database and revised the manuscript.
- Geneau G, Belisle A: Performed genotyping experiments.
- Provided studied samples: Garg AX, Clark WF, Moayyedi P, Collins SM, Marshall JK.
- Marshall JK and Franchimont D: Study conception, supervision and manuscript revision.

# CLAIMS FOR ORIGINALITY

1. *NLRP3* was identified as a novel Crohn's disease (CD) susceptibility gene.

2. We have identified 6 SNPs explaining the association signals, namely rs4353135, rs4266924, rs55646866, rs6672995, ss107635144 (novel SNP that we uncovered), and rs10733113.

3. We demonstrated that the genotypes of the variant rs4353135, located in the *NLRP3* CD-associated region, are associated with *NLRP3* gene expression in two independent sample set of healthy individuals.

4. We demonstrated that the genotypes of the variant rs667299, also located in the *NLRP3* CD-associated region, are associated with level of IL-1$\beta$ secretion.

5. We report *NLRP3* gene expression to be significantly upregulated in colon tissues isolated from mice with TNBS-induced colitis, and in biopsies from individuals with Crohn's disease.

6. We found *MEFV* expression to be significantly increased in both TNBS and DSS colitis mice model, and report an increase in *MEFV* gene expression in inflamed colonic tissues from both CD and ulcerative colitis (UC) patients, correlating with severity of inflammation.

7. Unlike previous studies (Cattan et al. 2000; Fidder et al. 2002; Fidder et al. 2005; Karban et al. 2005) that have excluded the involvement of *MEFV* in CD pathogenesis by looking at specific rare FMF missense causative mutations clustered in exon 2 and 10 in a relatively small number of CD cases, we carried out a thorough genetic screening of the *MEFV* region in two large CD sample sets and we report no significant associations between common variants in the *MEFV* region and CD susceptibility.

8. We also evaluated the *MEFV* region in 3 independent UC samples and report no consistent significant associations between common variants in the *MEFV* region and UC susceptibility.

9. We report the first gene-gene interaction analyses between *MEFV* and *NLRP3*, and show no signification contribution of these interactions to CD and UC susceptibility.

10. We report the first genetic association study ever performed in the field of post-infectious irritable bowel syndrome (PI-IBS).

11. We identified 3 novel PI-IBS susceptibility genes: *TLR9*, *CDH1*, and *IL6*.

# ABBREVIATIONS

| | |
|---|---|
| AID | auto-inflammatory diseases |
| AJC | apical junctional complex |
| AP-1 | activator protein 1 |
| ASC | apoptosis-associated speck-like protein containing a CARD |
| ATP | adenosine triphosphate |
| BIR | baculovirus inhibitor of apoptosis protein repeat domain |
| CAPS | cryopyrin-associated periodic syndromes |
| CARD | caspase-recruitment domain |
| CARDINAL | CARD-inhibitor of NF-κB activating ligand |
| CATERPILLER | CARD (caspase-recruitment domain) transcription enhancer, R (purine)-binding, pyrin, lots of leucine repeats |
| CC | coiled-coil |
| CD | Crohn's disease |
| cDNA | complementary deoxyribonucleic acid (DNA) |
| CI | confidence interval |
| CIITA | major histocompatibility complex [MHC] class II transactivator |
| CNV | copy number variations |
| CPPD | calcium pyrophosphate dihydrate |
| Ct | cycle threshold |
| DAMP | danger-associated molecular patterns |
| DD | death domain |
| DED | death effector domain |
| DNA | deoxyribonucleic acid |
| DSS | Dextran Sulfate Sodium |
| EST | expressed sequence tags |
| FBS | fetal bovin serum |
| FCAS | familial cold auto-inflammatory syndrome |
| FCU | familial cold urticaria |
| FIIND | domain with function to find |
| FMF | familial Mediterranean fever |
| GE | gastroenteritis |
| GI | gastro-intestinal |
| GWA | genome wide association |
| GWAS | genome wide association study |
| HIDS | hyperimmunoglobulinemia D with periodic fever syndrome |
| HLA | human leukocyte antigens |
| HPFS | hereditary periodic fever syndrome |
| HWE | Hardy-Weinberg equilibrium |
| IBD | inflammatory bowel disease |
| IBS | irritable bowel syndrome |
| IKK | IκB kinase |
| IL | interleukin |
| IL-1RA | interleukin 1 receptor antagonist |

| | |
|---|---|
| IPAF | IL-1β-converting enzyme [ICE]-protease-activating factor |
| IRF1 | interferon regulatory factor-1 |
| IRG | interferon-inducible guanosine triphosphatases |
| LD | linkage disequilibrium |
| LPS | lipopolysaccharide |
| LRR | leucine rich repeats |
| LTA | lipoteichoic acid |
| MAF | minor allele frequency |
| MAGUK | membrane-associated guanylate kinase |
| MAL | MyD88-adaptor-like |
| MDP | muramyl dipeptide |
| MHC | major histocompatibility complex |
| mRNA | messenger ribonucleic acid (RNA) |
| MSU | monosodium urate |
| MWS | Muckle-Wells syndrome |
| NACHT | domain present in Naip, CIITA, HET-E (plant het product involved in vegetative incompatibility) and TP-1 (telomerase-associated protein 1) |
| NAD | NACHT associated domain |
| NAIP | neuronal apoptosis inhibitor factors |
| NALP | NACHT domain, leucine-rich repeat and pyrin domains-containing protein |
| NBS | nucleotide binding site |
| NF-κB | nuclear factor-κB |
| NLR | Nucleotide oligomerization domain (NOD)-like receptor |
| NLRP3 | nucleotide binding oligomerization domain, leucine-rich-repeat family, pyrin containing 3 |
| NOD | nucleotide-binding oligomerization domain |
| NOD1 | nucleotide-binding oligomerization domain containing 1 |
| NOD2 | nucleotide-binding oligomerization domain containing 2 |
| NOMID | neonatal-onset multisystem inflammatory disease |
| OCTN | organic cation transporters |
| OMIM | Online Mendelian Inheritance in Man |
| PAMPs | pathogen-associated molecular patterns |
| PAPA | pyogenic sterile arthritis with pyoderma gangrenosum and acne |
| PBC | peripheral blood cells |
| PGN | peptidoglycan |
| PI-IBS | post-infectious irritable bowel syndrome |
| PRMs | pattern recognition molecules |
| PRRs | pattern recognition receptors |
| PYD | pyrin domain |
| RLR | Retinoic acid–inducible gene (RIG)-I-like receptor |
| RNA | ribonucleic acid |
| SD | standard deviation |
| SNP | single nucleotide polymorphism |
| Th1 | T helper 1 |
| Th17 | T helper 17 |

| | |
|---|---|
| TIR | Toll/IL-1 receptor |
| TJ | tight junctions |
| TLR | Toll-like receptor |
| TNBS | trinitrobenzene sulfonate |
| TNF | tumor necrosis factor |
| TRAM | TRIF-related adaptor molecule |
| TRAPS | tumor necrosis factor (TNF) receptor-associated periodic syndrome |
| TRIF | TIR-domain-containing adaptor protein inducing INF$\beta$ |
| UC | ulcerative colitis |
| WHS | Walkerton Health Study |
| WTCCC | Welcome Trust Case Control Consortium |

# 1 GENERAL INTRODUCTION

In the context of the 3 main projects presented in this thesis, a candidate gene approach was used to study three different disorders involving chronic intestinal inflammation that can result from a breakdown of the normal symbiosis between the mucosal immune system and the commensal flora, namely Crohn's disease (CD) and ulcerative colitis (UC) (together referred to as inflammatory bowel disease (IBD)), as well as post-infectious irritable bowel syndrome (PI-IBS) (Bouma et al. 2003; MacDonald et al. 2005; Karin et al. 2006; Cario 2008).

Indeed, these diseases appear to results from alterations of the microbial flora and dysregulation of the balance between tolerance and immunity towards this flora, leading to the activation of the intestinal innate immune system (MacDonald et al. 2005; Karin et al. 2006; Cario 2008). Although IBS is considered as a gut motility/functional disorder, the observation of persistent low-grade colonic inflammation in patients who develop IBS following gastroenteritis (namely PI-IBS) led to the speculation that PI-IBS could be part of the evolving IBD spectrum of inflammation, with CD and UC (i.e. IBD) at one end and normality at the other end (**Figure 1.1**) (Bradesi et al. 2003). This latter hypothesis is further supported by the fact that some PI-IBS patients may develop IBD (Quigley 2005; Bercik et al. 2005; Talley 2006; Burgmann et al. 2006).

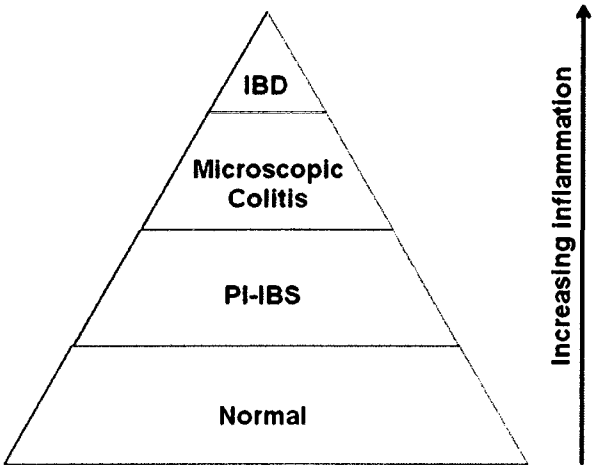



**Figure 1.1: Pyramid of gastro-intestinal disorders showing the spectrum of colonic inflammation**

PI-IBS: post-infectious irritable bowel syndrome; IBD: inflammatory bowel disease

This general introduction begins with the primary research objectives of the 4 Chapters of results presented in this thesis. Subsequent sections in this introduction outline key areas of the IBD pathophysiology and genetics, the candidate genes that have been selected, followed by the PI-IBS pathophysiology and an overview of the main axes of research in which the PI-IBS candidate genes were selected.

## 1.1 Overview of objectives

**OBJECTIVE #1:** To evaluate the contribution of the intracellular bacterial sensor NALP3 (*NLRP3* gene) to CD susceptibility:

This objective was reached by:

- Screening tagging single nucleotide polymorphisms (SNPs), sequencing the associated region, and performing thorough fine-mapping of the *NLRP3* region in an exploratory cohort of 296 father-mother-affected child CD trios.
- Replicating the association signals in 4 additional cohorts totaling 414 CD trios, 239 CD cases, and 107 healthy controls.
- Assessing *NLRP3* gene expression in experimental colitis mice models and CD patients.
- Evaluating the associations between SNP genotype and gene expression to determine whether this reflected *cis*-acting regulatory effects on *NLRP3* in two independent sample sets of healthy individuals.
- Evaluating whether SNPs would be associated with different levels of IL-1$\beta$ production, as NALP3 is involved in IL-1$\beta$ processing.

**OBJECTIVE #2:** To evaluate the contribution of the intracellular bacterial sensor pyrin (*MEFV* gene) to CD and UC susceptibility:

This objective was reached by:

- Assessing pyrin gene expression in experimental colitis mouse models and in CD and UC patients.

- Screening tagging SNPs and performing thorough fine-mapping of the *MEFV* region in an exploratory cohort of 512 CD trios, 239 CD cases, 113 UC trios, 96 UC cases, and 107 healthy controls.
- Sequencing and subsequent genotyping of variants in exonic and promoter regions located in the associated haplotype block.
- Attempting to replicate the association signals in 2 additional sample sets totaling 226 father-mother-affected child CD trios, 89 UC trios, 521 UC cases, and 300 healthy controls.

**OBJECTIVE #3:** To evaluate whether epistatic gene-gene interactions between *NLRP3* and *MEFV* contribute to CD and UC susceptibility:

This objective was reached by:

- Genotyping the 6 CD *NLRP3*-associated SNPs in the 202 UC trios, 96 UC trios, and 107 healthy controls used for the *MEFV* study.
- Stratifying analyses of *MEFV* variants conditional on each individual's genotypes at *NLRP3* variants using the CD or UC samples commonly genotyped in the *NLRP3* and *MEFV* studies.

**OBJECTIVE #4:** To identify genetic risk factors contributing to PI-IBS susceptibility:

This objective was reached by:

- Performing thorough literature review and identifying candidate functional variants involved in the serotoninergic, intestinal epithelial barrier and innate immune response.
- Evaluating 79 functional variants in the well-defined Walkerton health study cohort composed of Walkerton residents who developed PI-IBS after exposure to *Escherichia coli* 0157:H7 and *Campylobacter jejuni* in contaminated municipal water (Marshall et al. 2006).
- Performing fine-mapping in *TLR9*, *IL6*, and *CDH1* associated regions to confirm associations with PI-IBS.

## 1.2 Innate vs. adaptive immune response

The three projects presented in this thesis are focused primarily on candidate genes playing central roles in the intestinal innate immune response to pathogenic challenges in the gastro-intestinal (GI) tract. It is thus important to initially define the roles and differences between the innate and adaptive immune response systems.

Vertebrates are frequently challenged by a wide variety of microorganisms, predominantly bacteria and viruses, and throughout evolution, they have developed efficient strategies to fight off infections (Janeway et al. 2002).

Briefly, the immune response is mediated by two complementary systems: the innate and the adaptive immune systems.

The innate immunity is a phylogenetically ancient system and is generally directed towards immediate non-specific responses to commonly encountered threats in the environment (Janeway et al. 2002). Invasive bacterial pathogens use specific virulence factors that mediate efficient entry into the mammalian cells where they can evade the host immune mechanisms, such as complement- and antibody-mediated killing, and establish a replicative niche (Raskin et al. Cell 2006; Pizarro-Cerdá et al. 2006). This "strategy" ensures that, in most cases, the pathogen proliferates and successfully spreads to a new host. Meanwhile, because it does not require priming or the existence of cells with memory of earlier encounters with pathogens, the host innate immune system can provide a rapid response that serves multiple functions such as pathogen detection, antimicrobial defense, and instruction of an appropriate adaptive immune response (Delbridge et al. 2007). Such immediate activation of the host innate immunity relies on the host detection of conserved microbial motifs known as pathogen-associated molecular patterns (PAMPs), which are usually vital for microbial survival and, therefore, unlikely to vary in their structures (Akira et al. 2006). Examples of such molecules include parts of the bacterial cell envelope, like lipopolysaccharide (LPS) and peptidoglycan (PGN), flagellin and microbial nucleic acids (i.e. viral RNA or bacterial DNA) (Akira et al. 2006). These PAMPs are recognized by a limited number of germ-line-encoded receptors termed pattern recognition molecules (PRMs). This is unlike the case for the adaptive immune response, which requires somatic gene rearrangement for the

generation of epitope-specific antibodies (Palm et al. 2009). Control of infection does not always require adaptive immunity responses. Nevertheless, when necessary, the innate immune system is also very efficient in instructing the cellular-mediators of the adaptive immune system to generate additional "fighting" forces against the invading pathogens (Palm et al. 2009).

In contrast, the adaptive immune response is a highly specific response system that emerged more recently in the evolutions of vertebrate (Palm et al. 2009). It is concerned primarily with the development of long-lasting defense and memory of threats that have been encountered repeatedly. The ability of the adaptive immune system to adjust and to refine its repertoire over time based on the pathogens that it encounters results in a high level of efficiency. Its mechanisms involve the engagement of receptors that are selected for their reactivity to a specific antigen (i.e. T cells receptors and immunoglobin receptors on B cells) (Palm et al. 2009). The full development of these responses requires the expansion and differentiation of the specific responder cells, which have the role of establishing a memory for a specific antigen response. Cumulative evidence has shown that both the innate and adaptive immune systems appear to interact to form an efficient overall system of defense (Janeway et al. 2002; Palm et al. 2009).

# 1.3 Inflammatory bowel disease (IBD) overview

## 1.3.1 IBD clinical description

Crohn's disease (CD) and ulcerative colitis (UC), the two main types of inflammatory bowel disease (IBD), are chronic, relapsing inflammatory disorders of the GI tract that have a peak age of onset in the second to fourth decade of life. CD was first seen by German surgeon Wilhelm Fabry in 1623 (Fabry 1964), and was later described by and named after the US physician Burril B Crohn (Crohn et al. 1984). UC was first described by the British physician Sir Samuel Wilks in 1859 (Wilks 1859).

IBD patients typically suffer from frequent and chronically relapsing flares resulting in diarrhea, abdominal pain, rectal bleeding and malnutrition (Xavier et al. 2007). CD can be distinguished from UC in that the inflammation associated with CD is transmural (Xavier et al. 2007). The inflammation is often discontinuous, patchy and segmental, and it can affect the patient anywhere in the GI tract from the mouth to the anus (Xavier et al. 2007). However, CD most commonly involves the ileum and the colon and by contrast, the inflammatory changes in UC typically involve only superficial mucosal and submucosal layers of the intestinal wall (Xavier et al. 2007). The diffuse inflammatory pattern usually extends proximally from the rectum up to varying degrees in the colon. The main differences in inflammatory patterns between CD and UC are illustrated in **Figure 1.2**.

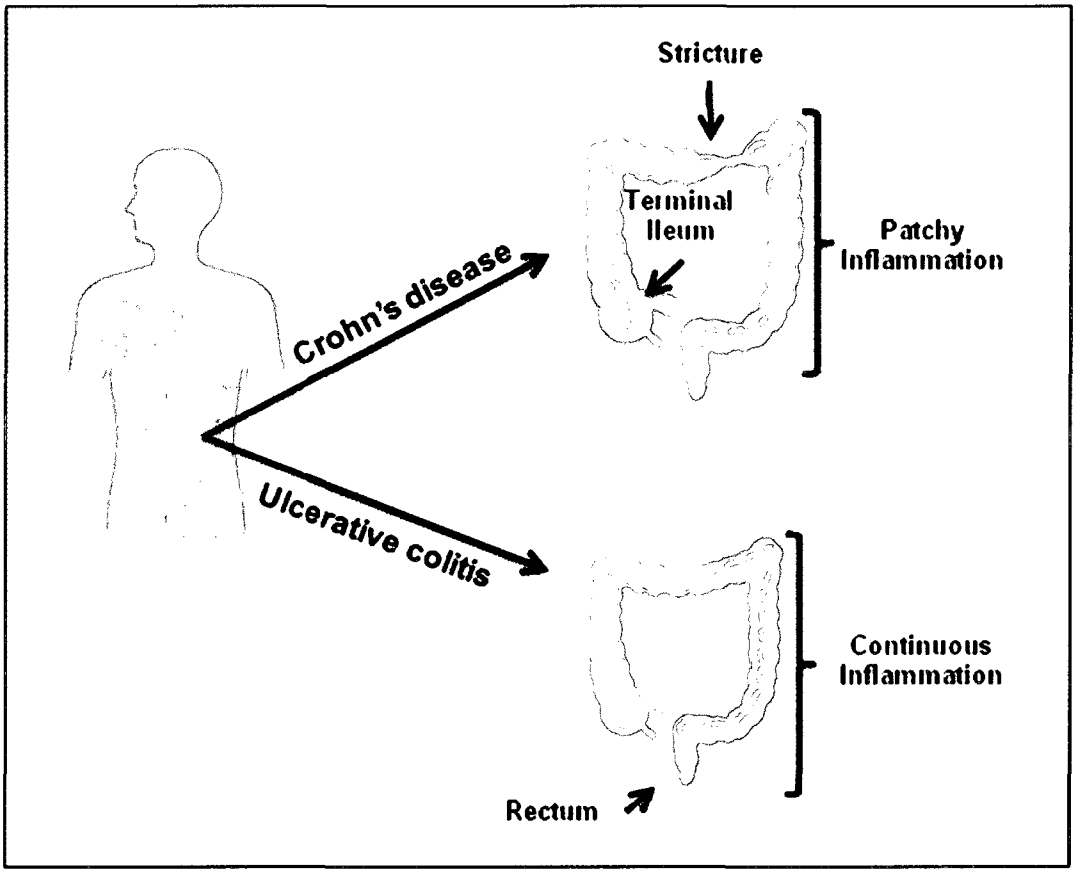



**Figure 1.2: Inflammatory distribution patterns of Crohn's disease and ulcerative colitis.**

Figure modified from (http://www.humanillnesses.com/original/Her-Kid/Inflammatory-Bowel-Disease.html)

Patients with IBD often suffer from various extra-intestinal manifestations, and are more likely to be affected by other chronic inflammatory diseases, particularly with primary sclerosing cholangitis, ankylosing spondylitis, and psoriasis (Loftus 2004).

The etiology of CD and UC is complex and not fully understood. The current understanding of these two disorders is that it appears to result from a dysregulated mucosal immune response to gut lumen bacterial antigens in a genetically susceptible host. In the following sections, each sphere (i.e. environmental, genetic, and immunobiology) contributing to IBD pathogenesis will be discussed.

### 1.3.2 CD & UC incidence

The highest incidence rates and prevalence of CD and UC have been reported from northern Europe, the UK, and North America, where the rates are beginning to stabilize (Baumgart et al. 2007). However, rates continue to rise in low-incidence areas such as southern Europe and Asia (Loftus 2004; Baumgart et al. 2007).

The combined prevalence of CD and UC in developed countries is estimated at 100-200 per 100,000 individuals (Podolsky 2002). In North America, the prevalence of CD in Caucasian individuals is estimated at 43.6 per 100,000 individuals (Kurata et al. 1992; Baumgart et al. 2007). Interestingly, other ethnicities in the same geographic area have a much lower CD prevalence, with 5.6 per 100,000 in Asians and 4.1 per 100,000 in Hispanic individuals (Kurata et al. 1992; Baumgart et al. 2007). These observations are in line with studies showing racial differences in disease location and extra-intestinal disease complications (Nguyen et al. 2006).

### 1.3.3 IBD environmental risk factors

#### 1.3.3.1 The acquisition, composition & role of the intestinal microbial flora

The bowel is the largest immunological organ of the body that is continuously exposed to interactions between the mucosal immune system and the intestinal microbial flora. It has been reported that an aberrant mucosal immune response to this flora may be involved in the initiation and perpetuation of IBD (Sartor 2008; Xavier et al. 2007). As such, the acquisition and composition of the intestinal microbial flora is one of the most important environmental factors influencing the development of this disorder.

The commensal bacteria community forms one of the most densely populated microbial habitats known in biology (Ley et al. 2006; Gill et al. 2006). Indeed, the mammalian intestine with its stable temperature and rich supply of nutrients provides a hospitable microenvironment for bacteria. The bacterial-derived signals from this flora are essential for normal intestinal physiology (Packey et al. 2008; Artis 2008). For example, commensal bacteria are known to facilitate digestion, absorption and storage of nutrients that would otherwise be inaccessible to the mammalian host (Stappenbeck et al. 2002; Guarner et al. 2003; Bäckhed et al. 2004; Rakoff-Nahoum et al. 2004; MacDonald et al. 2005b). Additionally, acquisition of commensal bacteria promotes the development of the intestinal epithelium, angiogenesis, and protection against tissue injury (Hooper et al. 2001; Rakoff-Nahoum et al. 2004). By occupying this environmental niche, the colonization of the intestine by the beneficial bacterial communities compete

for nutrients and secretion of antimicrobial peptides, which may provide a degree of protection for the host against the rapid colonization by pathogenic microorganisms (Sonnenburg et al. 2004; Bäckhed et al. 2005).

The colonization by commensal bacteria begins at birth with the initial exposure to the flora of the birth canal, and the bacterial communities are reported to reach a density of $10^{12}$ per ml of luminal contents in the adult human large intestine (Savage 1977; Stark et al. 1982; Benno et al. 1984). This acquisition appears to be the result of opportunistic colonization by particular species as a result of random environmental encounters (Palmer et al. 2007). Environmental and genetic factors have a major influence in the initial composition of commensal bacterial communities, and although a remarkable degree of inter-individual variation has been reported, the composition of the flora has been shown to be relatively stable over time in adults (Zoetendal et al. 1998; Eckburg et al. 2005; Palmer et al. 2007). These latter observations may explain why geographical and temporal variability, as well as lifestyle factors have been associated with the development of IBD (discussed bellow) since all of these variables may influence these random environmental encounters that participate in elaborating the initial composition of the microbial flora.

#### 1.3.3.2 Geographic, temporal and lifestyle influences

Several groups have reported a geographical gradient existing from Northern to Southern hemispheres correlating with variation in IBD incidence. This is illustrated by the fact that the highest reported incidence rates and

prevalence for CD and UC are from North America and northern Europe, and the lowest reported incidences rates are from South America, southeast Asia, Africa (with the exception of South Africa), and Australia (Loftus 2004; Economou et al. 2008). These observations could also indicate variation in access to, and quality of health care system, as well as differences in industrialization, sanitation, and hygiene. Prevalence has been reported to equilibrate over time with migration to other geographic areas, since immigrants moving from low incidence regions to developed countries were reported to have an increase in incidence rates of IBD (Farrokhyar et al. 2001; Zheng et al. 2005; Baumgart et al. 2007; Economou et al. 2008). Similar to the gradient difference mentioned above, epidemiological studies conducted in North America and Europe have also reported an accumulation of IBD cases in urban compared to rural communities (Farrokhyar et al. 2001; Economou et al. 2008; Aamodt et al. 2008). Again, these results could be explained by different lifestyle habits related to sanitation and hygiene, or differences in access to specialized health care facilities.

Interestingly, several countries with historically lower incidence rates of IBD have shown in the past two decades, a pattern of rising incidence, suggesting that environmental risk factors are also involved (Baumgart et al. 2007; Economou et al. 2008; de Mesquita et al. 2008). Some of these factors may come from the individual's lifestyle, which can also influence the composition of the intestinal microbial flora. Indeed, several studies have reported associations between breastfeeding, hygiene, infection, psychological stress, and appendectomy with susceptibility to developing IBD (Baumgart et al. 2007; de Mesquita et al. 2008).

# 1.4 IBD Genetic risk factors

## 1.4.1 Twin study

In recent years, there has been accumulating evidence supporting the involvement of genetic risk factors contributing to IBD pathogenesis. The strongest evidence comes from concordance studies in twins, which represent a specific type of family study that provides a unique opportunity to discern the genetic and environmental contributors to the disease (Hawkes 1997). Because monozygotic twins share the exact same genetic code and dizygotic twins share half of the same genetic code, comparing the concordance rates between monozygotic and dizygotic twins allows the quantifying of the genetic and environmental contributions to disease development (Hawkes 1997). The first reported concordance analysis result (Tysk et al. 1988) was subsequently supported by two other studies (Thompson et al. 1996; Orholm et al. 2000), showing a pooled concordance in monozygotic twins of 37.3% for CD and 10% for UC (Baumgart et al. 2007). In parallel, pooled concordance for dizygotic twins has been reported to be of 7% for CD and 3% for UC (Tysk et al. 1988; Thompson et al. 1996; Orholm et al. 2000; Baumgart et al. 2007). These observations suggested that the genetic contribution in CD was more important than that for UC.

### 1.4.2 Familial aggregation & first degree relatives

Other evidence for the genetic involvement in IBD includes familial aggregation of the disease (Russell et al. 2004). A positive family history is still considered as one of the largest independent risk factors for IBD; first-degree relatives of affected individuals have a relative risk of five-fold or more of developing IBD (Orholm et al. 1991; Russell et al. 2004; Baumgart et al. 2007). People with CD have a first-degree relative with CD in 2.2-16.2% of cases and with IBD in 5.2-22.5% of cases (Russell et al. 2004; Baumgart et al. 2007). In the case of UC, patients have a first degree relative with UC in 5.7-15.5% of cases and with IBD in 6.6-15.8% of cases (Russell et al. 2004; Baumgart et al. 2007). These results further support the observation that the inheritance component seems to be stronger in CD than in UC (Tysk et al. 1988; Orholm et al. 1991).

### 1.4.3 Ethicity

In addition, ethnical aggregation has also been observed in certain communities. For example, CD is known to be more prevalent in Jewish people than in any other ethnic group, a phenomenon known as genetic anticipation (Heresbach et al. 1998; Hampe et al. 2000). UC is three to five times more prevalent in Jewish people, and important epidemiological differences exist between Jewish people living in Israel and those living elsewhere (Niv et al. 2000). In Israel, Ashkenazi Jews have a higher incidence than Sephardi Jews, but a lower incidence than Ashkenazi Jews living in the United States or Northern Europe.

### 1.4.4 IBD susceptibility genes: candidate vs. genome wide association study (GWAS) approaches

CD and UC are polygenic heterogeneous diseases. It is now known that multiple genes are involved, all conferring a small increase in disease risk, and no single susceptibility gene is either necessary or sufficient to produce IBD (Cho 2008). To help further elucidate the key pathogenic pathways involved in CD and UC, several groups have undertaken genetic association studies. Although compelling evidence from epidemiological studies supporting the role of genetic risk factors in IBD has been available for decades, the identification of the genes involved has proven rather challenging. Indeed, progress in delineating the IBD genetic architecture up until 2006 has been very slow, but recent findings have helped shaped the basis of IBD immunopathogenesis and will be reviewed below.

Broadly, two different strategies have been used when undertaking these genetic studies. The first one is a hypothesis-based approach focusing on specific candidate genes, and the other one is based on hypothesis-free methods and include linkage and genome-wide association studies (GWAS).

#### 1.4.4.1 Hypothesis driven approach: candidate gene

The candidate gene approach is based on the specific attributes of the gene and its encoded protein that could potentially implicate it in the disease pathogenesis. These include expression or functional data, results from linkage studies, animal models, or known associations in other related diseases. This very intuitive approach has had very little success in uncovering novel IBD

susceptibility genes, perhaps due to the poor study designs that have been previously used (Cardon et al. 2001).

#### 1.4.4.2 Hypothesis-free approach: GWAS and linkage

The GWAS hypothesis-free method has evolved from the low-resolution linkage analysis, a method based on multiple affected families that was widely used in the late 1990s. GWAS involves genotyping a large number of SNPs that sample human genetic variations throughout the genome. Several key tools have advanced the development of the GWAS platforms. These include 1) the discovery of more than 10 millions SNPs in the human genome (i.e. dbSNP database), which is a large resource of variants that can be used to evaluate disease association (Sherry et al. 1999; Phillips 2007); 2) the development of a high-resolution genetic map of the human genome by the HapMap project that has allowed to measure the association between alleles of neighboring SNPs (i.e. linkage disequilibrium (LD)) and helped to define haplotypes (International HapMap Consortium 2005; International HapMap Consortium et al. 2007; Manolio et al. 2008); and 3) the development of cutting-edge genotyping platforms that are now capable of analyzing up to 1 million SNPs simultaneously on a single array and at relatively feasible cost (Grant et al. 2008; Beaudet et al. 2008). Indeed, the HapMap database (Manolio et al. 2008) can now be used to design genotyping platforms that can efficiently assay a large fraction of common genetic variants with a minor allele frequency (MAF) greater than 5%. The GWAS approach has addressed and resolved 1) the poor resolution issues of

linkage analysis by offering a much greater coverage of the genome and 2) the poor reproducibility issues of many candidate gene studies by the development and use of stringent statistical thresholds to explicitly account for genome-wide multiple testing and limit the generation of false positive results (Hirschhorn et al. 2005). The GWAS strategy has the advantage of offering an unbiased survey of the whole genome searching for IBD-associated loci, and it can provide novel insights on disease pathways and mechanisms involved in the origin and development of IBD.

### 1.4.5 The IBD "historical" associations

Until two years ago with the publication of several GWAS, 15 years of endeavor by multiple groups had resulted in the identification of only a few reproducible loci. In 2001, a linkage study and a positional cloning study independently led to the identification of the first CD susceptibility gene, *NOD2* (Hugot et al. 2001; Ogura et al. 2001), which is a pattern-recognition receptor playing a key role in intracellular bacterial sensing and initiation of the innate immune response. Linkage analysis and regional association studies have also identified susceptibility loci in a broad region on chromosome 5 (also referred to IBD5) that contains several potential candidate genes having roles in the inflammatory process (Rioux et al. 2001; Peltekova et al. 2004) and in the 10q23 chromosomal region (disc large homologue 5 (DLG5)) (Stoll et al. 2004), with strong and less consistent replication for IBD5 and DLG5, respectively. Genome-wide linkage studies of CD conducted over the past 15 years have implicated

many other loci on 12 different chromosomes (Gaya et al. 2006). According to their initial date of reporting, the regions on chromosomes 16, 12, 6, 14, 5, 19, 1, 16, and 3 have been renamed as IBD1-9, respectively (Baumgart et al. 2007). No single locus was consistently reported in all these genome scans. The most-well replicated of these loci is IBD3 located on chromosome 6p, where the major histocompatibility complex (MHC) lies (Baumgart et al. 2007). The status for the other loci is less clear. Between 2001-2006, several other groups have reported potential IBD candidate genes using either linkage or candidate gene approaches (Franchimont et al. 2004; Low et al. 2004; McGovern et al. 2005; van Bodegraven et al. 2006; Annese et al. 2006; De Jager et al. 2007), but the results of the vast majority of these studies have not resisted to rigorous analyses and have failed to replicate subsequently (Cardon et al. 2001).

#### 1.4.5.1 NOD2

##### *NOD2 CD associated mutations*

As mentioned previously, two independent studies, one using linkage analysis and one using positional cloning, led to the identification of the nucleotide-binding oligomerization domain containing 2 (*NOD2*; also known as *CARD15*), the first CD susceptibility gene (Hugot et al. 2001; Ongura et al. 2001) and perhaps the CD gene that is the most characterized so far. *NOD2* encodes for a protein comprising two caspase-recruitment domains (CARDs), a nucleotide binding site (NBS), and multiple leucine rich repeats (LRR), which recognizes PAMPs (Franchi et al. 2009). Three rare mutations in *NOD2* with a MAF lower

than 5% in healthy individuals were each associated with CD susceptibility. Two of the mutant alleles, Arg702Trp and Gly908Arg, result in single amino acid substitutions within the LRR, whereas the third, Leu1007fsinsC, lead to a frame-shift mutation that removes the last 33 amino acids of the NOD2 polypeptide (Hugot et al. 2001). These findings have been widely replicated across several cohorts of European ancestry, but these mutations appear to be absent in CD patients of Asian ancestry (Yamazaki et al. 2002; Croucher et al. 2003; Leong et al. 2003) and are very rare in patients of African descent (Kugathasan et al. 2005).

It has been estimated that mutations in *NOD2* are present in up to one-third of individuals with CD (Hugot et al. 2001; Ogura et al. 2001). A meta-analysis of 39 studies, all looking at samples from European descent, has estimated an odds ratio for a simple *NOD2* carrier (i.e. heterozygote) of 2.4 (confidence interval (CI): 2.0-2.9) and of 17.1 (CI: 10.7-27.2) for *NOD2* homozygotes or compound heterozygotes (Economou et al. 2004). Despite the high odds ratio associated with these mutations, the disease penetrance is estimated to be very limited, as many patients who are monozygotes for these mutations remain healthy, and even those who develop the disease are commonly asymptomatic for the first 10-15 years of their life (Economou et al. 2004; Hugot et al. 2007). These latter observations suggested that additional genetic and environmental factors acted in conjunction with the NOD2 variants in contributing to CD development (Hugot et al. 2001; Ogura et al. 2001; Economou et al. 2004; Hugot et al. 2007). Further supporting this latter hypothesis is the fact that NOD2-deficient mice do not show evidence of intestinal inflammation (Pauleau et al. 2003; Kobayashi et al. 2005).

### *NOD2: a PRM for PGN and MDP*

NOD2 is an intracellular pattern-recognition molecule (PRM) predominantly expressed in monocytes and Paneth cells that senses for bacterial peptidoglycan (PGN) and that can be activated by muramyl dipeptide (MDP), the minimal bioactive component of PGN (Girardin et al. 2003; Inhora et al. 2003). PGN is present in both Gram-positive and Gram-negative bacteria, is responsible for providing shape and mechanical rigidity to bacteria, and is composed of glycan chains that contain alternating *N*-acetylmuramic acid (MurNAc) and *N*-acetylglucosamine (GlcNAc) sugars cross-linked to each other by short peptides (Carneiro et al. 2007). During the bacterial life cycle, PGN is constantly degraded by specific hydrolases as newly synthesized subunits are integrated into the polymeric structure to allow different processes such as cell division (Carneiro et al. 2007). The PGN degradation leads to the formation of muropetides, such as MDP, that have strong immunomodulary properties and that are recognized by NOD2 (Girardin et al. 2003; Boneca 2005).

### *NOD2: initiation of innate immune response*

*NOD2* is thus known to play a key role in initiating the innate immune response to bacterial bi-product. The consequence of NOD2 activation is thought to largely result in the activation of nuclear factor-κB (NF-κB) (Inohara et al. 2003). NF-κB is a heterodimeric transcription factor and crucial component of the inflammatory response. It is involved in the regulation of genes encoding for proinflammatory cytokines, adhesion molecules, chemokines, growth factors, and

inducible enzymes such as cyclooxygenase 2 and inducible nitric oxide synthase (Baldwin 2001; Carneiro et al. 2007). The production of these inflammatory mediators is needed to facilitate clearance of bacteria (Kelsall 2005).

PGN or MDP sensing is thus followed by oligomerization of NOD2 protein, which in turn recruits a serine-threonine kinase called Rip2 (also known as RICK or CARDIAK), through homotypic CARD-CARD interactions (Carneiro et al. 2008). Rip2 then interacts with IKKγ (also known as NEMO), the regulatory subunit of the IκB kinase (IKK) complex, promoting its ubiquitination and subsequent degradation of IkBα, which in steady state condition retains NF-κB in the cytosol (Carneiro et al. 2008). Once degraded, NF-κB is released and translocates into the nucleus, where it drives the transcription of target genes (Burns et al. 2004).

### *Functional consequences of carrying NOD2 mutations*

The CD-associated *NOD2* mutations are located in or close to the C-terminal LRR region of NOD2 (Hugot et al. 2001; Ogura et al. 2001). The CD-associated mutations are thought to interfere with the ability of NOD2 to recognize the ligand, resulting in a reduced capacity to activate NF-κB in response to MDP stimulation (Cho 2008). Yet, the precise mechanism by which *NOD2* mutations contribute to CD susceptibility remains controversial as multiple alterations of NOD2 function have been reported and proposed. Van Heel *et al.* performed detailed functional analyses in human peripheral monocytes and reported that CD-associated NOD2 mutations caused a "loss of function"

phenotype characterized by impaired interleukin (IL)-1$\beta$ secretion upon stimultation with MDP or PGN (van Heel et al. 2006). Additional studies using cells derived from CD patients also showed defective NOD2 signaling in monocytes carrying the mutations, characterized by diminished MDP-mediated secretion of IL-1$\beta$ or IL-8 (Li et al. 2004; van Heel et al. 2005). Finally, an inappropriate immune response has also been reported in MDP stimulated dendritic cells from CD patients harboring NOD2 mutations, which was characterized by suppression of the anti-inflammatory cytokine IL-10 secretion, suggesting a loss of regulatory elements required to contain the inflammatory response (Kramer et al. 2006).

In a study where Watanabe *et al.* looked at determining the conditions leading to the development of colitis in mice with NOD2 deficiency, they report that colitis was highly dependent on TLR2 function since it was suppressed in NOD2 and TLR2 double-deficient mice (Watanabe et al. 2006). They suggested that NOD2-deficient mice become susceptible to colitis as a result of increased TLR2 responses when they have the capacity to respond to an antigen expressed by mucosal bacteria (Watanabe et al. 2006). Additional research is needed to evaluate whether NOD2 cross-talk with other PRMs.

In contrast to the loss-of-function mechanisms suggested above, Maeda *et al.* suggested a "gain-of-function" scenario associated with the NOD2 variants (Maeda et al. 2005). This group generated a knockout-knockin *NOD2* mouse by introducing a mutation homologous to the human L1007fsinsC mutation. Stimulation of macrophages from this mouse with MDP or PGN resulted in a

more effective activation of RIP2 and NF-κB as well as an enhanced caspase-1 mediated IL-1β processing and release (Maeda et al. 2005). They suggested that the CD-associated NOD2 mutations resulted in a "gain-of-function phenotype", and that the enhanced IL-1β release from the macrophages could contribute to the intestinal inflammation through the activation of a broader range of pro-inflammtory genes in an autocrine or paracrine manner (Maeda et al. 2005). Given that these latter results are in contrast to the reports demonstrating blunted responses to MDP or PGN in human cells expressing the NOD2 mutations, further species-specific and cell type-specific research is still needed to elucidate the function of the NOD2 CD-associated mutations.

### *NOD2: an anti-bacterial factor*

NOD2 has also been proposed to act as an anti-bacterial factor in human epithelial cells. Indeed, the expression of NOD2 frameshift mutant (Leu1007fsinsC) in intestinal epithelial cells resulted in an impaired capacity of these cells to control *Salmonella typhimurium* intracellular growth (Hisamatsu et al. 2003). NOD2 is also expressed in Paneth cells, which are located in the epithelial crypts along the GI tract (Elphick et al. 2005; Wehkamp et al. 2006). These cells provide protection against intraluminal bacteria by secreting anti-bacterial substances such as defensins (Elphick et al. 2005; Wehkamp et al. 2006). Kobayashi *et al.* reported that NOD2-deficient mice exhibit decreased defensin expression that is associated with an inability to detect MDP, and an increased susceptibility to oral infection with *Listeria monocytogenes* (Kobayashi et al.

2005). It has also been reported that CD patients carrying NOD2 mutations have decreased expression of human defensins 5 and 6 in affected ileum tissues (Wehkamp et al. 2005). This same study showed that CD patients bearing NOD2 mutations have a significant decrease in α-defensin expression (Wehkamp et al. 2005). Such decrease in expression may contribute to the impairment of microbial clearance at mucosal surfaces and could partially account for the increased colonization of entero-adherent *Escherichia coli* observed in the ileal mucosa of some CD patients (Darfeuille-Michaud et al. 2004). Overall, these observations support a "loss-of-function" scenario and suggest that, in normal conditions, NOD2 expression in Paneth cells may contribute to epithelial host defense against bacterial insults through the induction of anti-bacterial peptides such as defensins; NOD2 variants might lead to failure to activate those defense mechanisms, resulting in increased bacterial burden followed by mucosal inflammation (Fritz et al. 2006; Franchi et al. 2009).

The precise role of NOD2 as a mediator of innate immunity remains poorly understood, but it is clearly a key sensor of intracellular bacterial MDP, and abnormalities in the handling of intracellular bacteria are emerging as a key theme in CD pathogenesis (Cho 2008).

#### 1.4.5.2 IBD5

The fist evidence for a susceptibility locus in the 5q31 region was obtained though linkage analysis of Canadian CD affected sibling pair families (Rioux et al. 2001). Fine mapping of this linkage region led to the identification of a 250-kb haplotype conferring susceptibility to CD (Rioux et al. 2001). However, the identification of the causal gene(s) and variant(s) responsible for the association signal has proven challenging because of the strong linkage disequilibrium (LD) in the region (Rioux et al. 2001; Reinhard et al. 2006). This region contains a cytokine gene cluster comprising several genes important in the maintenance of epithelial integrity and immunoregulation such as interferon regulatory factor-1 (*IRF1*), IL-4, IL-5, and IL-13, and organic cation transporters (OCTN) 1 and 2 (Reinhard et al. 2006).

Peltekova *et al.* reported a two-locus risk haplotype in the region of the OCTN genes and suggested that these accounted for the CD associations (Peltekova et al. 2004). More specifically, this group identified mutations within OCTN1 (*SLC22A4*; 1672C/T or L503F) and OCTN2 (*SLC22A5*; -207G/C) and reported an association between the haplotype comprising the risk alleles (-TC) and increased susceptibility to CD (Peltekova et al. 2004). Although a large number of replication studies have confirmed the association with the IBD5 locus, detailed analyses in the Canadian and other populations that included other polymorphisms in the IBD5 locus demonstrated that several other candidates within this cytokine gene cluster are as likely to be the causative genes as well,

thus leaving the question as to which is the causative IBD5 variant(s) unanswered (Silverberg 2006; Reinhard et al. 2006; Silverberg et al. 2007).

#### 1.4.5.3 DLG5

Located on chromosome 10q23, *DLG5* is the human homolog of the drosophila gene discs large homolog 5 (Stoll et al. 2004; Friedrichs et al. 2006). *DLG5* is a member of the membrane-associated guanylate kinase (MAGUK) family of scaffolding proteins known to be important in signal transduction and epithelial-cell integrity (Stoll et al. 2004; Friedrichs et al. 2006). Since the paper of Stoll *et al.* reporting DLG5 as a novel CD susceptibility locus (Stoll et al. 2004), there have been several additional studies evaluating variants of the DLG5 region in CD, but only a few reported replication and association with CD (Friedrichs et al. 2006). The exact reasons for the discrepancies amongst the studies remain unknown, but a recently published meta-analysis including most published studies suggested that DLG5 does not play a major role in CD susceptibility (Browning et al. 2007).

#### 1.4.5.4 MHC complex

The most convincing evidence for linkage to IBD across all populations according to two large meta-analyses, is located on chromosome 6 (IBD2), which encodes for the major histocompatibility complex (MHC) (Williams et al. 2002; van Heel et al. 2004). In addition to linkage studies, GWAS and candidate gene studies have also implicated variants located in the MHC complex in IBD

susceptibility, especially in UC (Satsangi et al. 1996; Stokkers et al. 1999; Franke et al. 2008a; Fisher et al. 2008; Kugathasan et al. 2008; Silverberg et al. 2009). This region comprises several immunoregulatory genes, including the human leukocyte antigens (*HLA*) genes and tumor necrosis factor alpha (TNF-$\alpha$/*TNF*), and has met the criteria of GWA significance in a published meta-analysis of approximately 2000 affected sibling pairs (van Heel et al. 2004). There is a possibility that more than one susceptibility locus within the larger HLA region contributes to IBD. However, the extensive LD across the region and the genetic heterogeneity between the different studied populations (i.e. European and Japanese) has limited the identification of the causal variant(s) (de Bakker et al; 2006). Despite the fact that the pro-inflammatory cytokine *TNF* could be a strong candidate for the region, numerous candidate gene association studies have failed to clearly demonstrating its implication in CD (Bouma et al. 1996; Cantor et al. 2005; Tremelling et al. 2006).

### 1.4.6 Genome wide association studies: bringing further insight into IBD pathogenesis

The introduction of GWAS has produced sea-change in complex disease genetics and has been particularly successful in CD, leading to the discovery of several new putative susceptibility genes. Seven independent CD GWAS with varying size and power were published between 2005-2007 (Yamazaki K et al. 2005; Duerr et al. 2006; Hampe et al. 2007; Rioux et al. 2007; Libioulle et al. 2007; Wellcome Trust Case Control Consortium 2007a; Parkes et al. 2007;

Franke et al. 2007; Raelson et al. 2007), together with one meta-analysis of three of these CD GWAS (Barrett et al. 2008), and one GWAS in IBD pediatric patients (Kugathasan et al. 2008). Additionally, three GWAS conducted specifically on UC were reported between 2008-2009 (Franke et al. 2008a; Fisher et al. 2008; Silverberg et al. 2009). The features and main results of these studies are summarized in **Tables 1.1** and **1.2**.

**Table 1.1: Compilation of the genome-wide association studies results**

| Population and Study | Disease | Sample size (cases-controls)[n] | Genotyping technology platforms | Chromosome | Genes or loci | Best associated SNPs | Location[o] |
|---|---|---|---|---|---|---|---|
| **Japanese[a]** | CD | 94-752 | 72K | 9q32 | TNFSF15 | rs6478108 | 116598524 |
| **German[b]** | CD | 735-368 | 7K (nsSNPs) | 2q37.1 | ATG16L1 | rs2241880 | 233848107 |
| **North American (European)[c,d]** | CD | 946-977 | 304K (Illumina HumanHap 300) | 1p31 | IL23R | rs7517847 | 67454257 |
| | | | | 2q37.1 | ATG16L1 | rs2241880 | 233848107 |
| | | | | 4p12 | PHOX2B | rs16853571 | 41447887 |
| | | | | 10q21 | None | rs224136 | 64140681 |
| | | | | 16q24.1 | FAM92B | rs8050910 | 83696674 |
| | | | | 22q13.1 | NCF4 | rs4821544 | 35588449 |
| **Belgian/French[e]** | CD | 547-928 | 302K (Illumina HumanHap 300) | 1p31 | IL23R | rs11209026 | 67478546 |
| | | | | 5p13.1 | Gene desert | rs1002922 | 40422312 |
| **United Kingdom[f,g]** | CD | 1748-2938 | 469K (Affymetrix 500K) | 1p31 | IL23R | rs11805303 | 67448104 |
| | | | | 2q37.1 | ATG16L1 | rs10210302 | 233823578 |
| | | | | 3p21 | MST1 | rs9858542 | 49676987 |
| | | | | 5p13.1 | Gene desert | rs9292777 | 40473705 |
| | | | | 5q33.1 | IRGM | rs13361189 | 150203580 |
| | | | | 10q21 | None | rs10761659 | 64115570 |
| | | | | 10q24.2 | NKX2-3 | rs10883365 | 101277754 |
| | | | | 18p11.3-p11.2 | PTPN2 | rs2542151 | 12769947 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **German**[h] | CD | 393-399 | 92K (Affymtrix 100K) | 11p15.2-p15.1 | NELL1 | rs1793004 | 20655505 |
| | | | | 5p13.1 | Gene desert | rs1992660 | 40450824 |
| **Quebec**[i] | CD | 382 trios | 164K (Perlegen) | 1p31 | IL23R | rs7518660 | 67458031 |
| | | | | 3p21.3 | Many | rs1131095 | 49689229 |
| | | | | 4p16.1 | JAKMIP1/LOC285484 | rs10003892 | 6257685 |
| | | | | 17q11.1 | Many | rs4435306 | 23573843 |
| | | | | 17q22-q23 | Many | rs6504016 | 56513025 |
| **German**[j] | UC | 1167-777 | 440K (Affymetric 5.0) | 1q32 | IL10 | rs3024505 | 205006527 |
| | | | | 2q35 | ARP2C | rs12612347 | 218765583 |
| | | | | 6p21 | HLA-DRA/HLA class II | rs9268877, rs9268480, rs9268858 | 32471822-32539125 |
| | | | | 1p31 | IL23R | rs11805303 | 67448104 |
| **United Kingdom**[k] | UC | 905-1465 | 10.8K (nsNPs) | 1q21.2 | ECM1 | rs3737240, rs13294 | 148749979-148751611 |
| | | | | 6p21 | HLA-DRA/HLA class II | rs6927022 | 32720375 |
| | | | | 3p21 | MST1 | rs3197999 | 49696536 |
| | | | | 1p31 | IL23R | rs11209026 | 67478546 |
| | | | | 5q31.1-q33.1 | IL12B | rs6556416 | 158751323 |
| | | | | 10q24.2 | NKX2-3 | rs10883365 | 101277754 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **North American (European)**[l] | UC | 1052-2571 | 300K and 500k (Illumina Human Hap 300v2 & HumanHap550v3) | 1p36 | PLA2G2E | rs6426833 | 20044447 |
| | | | | 1p36 | RNF186 | rs3806308 | 20015453 |
| | | | | 12q15 | IFNG/IL22 | rs1558744 | 66790859 |
| | | | | 12q15 | IL26 | rs2870946 | 66882928 |
| | | | | 1p31 | IL23R | rs11209026, rs10889677 | 67478546-67497708 |
| | | | | 6p21 | HLA-DRA/HLA class II | rs2395185 | 32541145 |
| **North American (European: pediatric-onset)**[m] | IBD | 1011-4250 | 517K (Illumina Human Hap550) | 20q13 | TNFRSF6B | rs2315008, rs4809330 | 61814400-61820030 |
| | | | | 21q22 | PSMG1 | rs2836878 | 39387404 |
| | | | | 6p21 | HLA-DRA/HLA class II | rs2395185, rs477515, rs2516049, rs9271568 | 32541145-32698441 |

[a] Yamazaki K et al. 2005; [b] Hampe et al. 2007; [c] Duerr et al. 2006; [d] Rioux et al. 2007; [e] Libioulle et al. 2007; [f] Wellcome Trust Case Control Consortium 2007a; [g] Parkes et al. 2007; [h] Franke et al. 2007; [i] Raelson et al. 2007; [j] Franke et al. 2008a; [k] Fisher et al. 2008; [l] Silverberg et al. 2009; [m] Kugathasan et al. 2008; [n] Sample numbers that were successfully genotyped; [o] Position according to dbSNP build 129, NCBI Build 36.1 (hg18).

**Table 1.2: Genome-wide association studies results: brief description of the associated gene function**

| Genes/ loci | Gene function | Disease[n] | Studies[o] |
|---|---|---|---|
| *TNFSF15* | Encodes a cytokine of the tumor necrosis factor (TNF) ligand family, which is induced by TNF and can activate NFκB (Yamazaki K et al. 2005) | CD | Japanese[a] |
| *ATG16L1* | Involved in autophagy, which removes intracellular pathogens and supports both the innate and acquired immune response (Xavier et al. 2008) | CD | German[b], North American[d], United Kingdom[f,g] |
| *PHOX2B* | Homeobox transcription factor involved in neural crest development. *PHOX2B* is mutted in hypoventilation syndrome, neuroblastoma and Hirschsprung disease (Cargnin et al. 2005; Pasqualetti et al. 2001). | CD | North American[c,d] |
| *FAM92B* | Predicted gene supported by multiple mRNAs, spliced ESTs and significant conservation across multiple species. It encodes a protein with no known function or recognizable motifs (Rioux et al. 2007) | CD | North American[c,d] |
| *NCF4* | The regulatory component of the superoxide-producing phagocyte NADPH-oxidase, which is involved in host defense (Ellson et al. 2006;Suh CI et al. 2006). | CD | North American[c,d] |
| Gene desert | Implicated in the regulation of prostaglandin E receptor 4 gene (PTGER4) (Libioulle et al. 2007) | CD | North American[c,d], United Kingdom[f,g], German[h] |
| *IRGM* | Involved in autophagy, which removes intracellular pathogens and supports both the innate and acquired immune response (Xavier RJ et al. 2008). | CD | United Kingdom[f,g] |
| *NKX2-3* | Homeobox transcription factor required for T- and B-cell localization in the spleen and mesenteric lymph nodes (Pabst et al. 2000). | CD | United Kingdom[f,g, k] |
| *PTPN2* | Encodes a T-cell protein tyrosine phosphatase and is a negative regulator of inflammatory response (Parkes et al. 2007) | CD | United Kingdom[f,g] |
| *NELL1* | Encodes a protein containing epidermal growth factor (EGF)-like repeats. Aberrant expression in mice is associated with defects in cranial and skeletal development (Franke et al. 2007) | CD | German[h] |
| *JAKMIP1* | JAKMIP1 is involved in IL23 signaling and binds to TYK2, a member of the Janus (tyrosine) kinase (Jak) family, which is associated with the IL12RB chain and mediates STAT-4 activation (Watford et al. 2004). | CD | Quebec[i] |
| *LOC285484* | LOC285484, contains a characteristic high-mobility group (HMG)-box domain (Yang D et al. 2007). | CD | Quebec[i] |
| *IL10* | *Il10*$^{-/-}$ mice develop spontaneous colitis phenotype, resulting from a defective anti-inflammatory counter-regulation in response to the commensal flora (Sellon et al. 1998). Therapeutic administration of human recombinant IL10 to individuals with UC had a positibe clinical effect (Shreiber et la. 1995; Fedorak et al. 2000) | UC | German[j] |
| *ARP2C* | The Arp2/3 protein complex has been implicated in the control of actin polymerization in cells. It interacts with protein encoded by *WAS* gene, which is involved in the regulation of T cells (Marangoni et al. 2007). | UC | German[j] |
| *HLA-DRA/ HLA class II* | Supports the large body of evidence suppoting an association between cliassicla HLA loci and UC (Stokkers et al. 1999). | UC | German[j], United Kingdom[k], North American[l,m] |
| *ECM1* | Encodes the extracellular matrix protein 1, a glycoprotein expressed in the small and large intestine, interacting with the basement membrane, inhibiting matrix metalloproteinase 9 (Chan et al.2007), and trongly activating NFκB signaling (Matsuda et al. 2003). | UC | United Kingdom[k] |

| | | | |
|---|---|---|---|
| *IL12B* | Encodes a subunit shared by IL-12 and IL-23. It is involved in the Th17 pathways and cell-mediated immunity against infectious pathogens (Abraham C amd Cho JH. Ann Rev Med 2008) | UC | United Kingdom[k] |
| *PLA2G2E* | Member of the secretory phospholipase A2 family of proteins that release arachidonic acid from membrane phospholimids, leading to the production of proinflammatory lipid mediators, such as prostaglandins and leukotrienes (Murakami et al. 2002). PLA2G2E expression in small intestine is induced by LPS, suggesting a role in bacterially associated inflammation (Suzuki et al. 2000). | UC | North American[l] |
| *RNF186* | Ring finger protein 186 (RNF186). Ring finger proteins are involved in ubiquitination of proteins and in diverse cellular processes (Joazeiro et al. 2000). | UC | North American[l] |
| *IFNG/IL22* | INF-g is critical in the immune response to pathogens, through regulation of macrophage function; it regulated many levels of immune homeostatsis, including T-cell subsets, NK cells and NKT cells (Schoenborn et al. 2007). IL-22 mediates host defense to bacterial pathogens (Aujla et al. 2007) as well as protection during acute inflammation, including intestinal inflammation (Sugimoto et al. 2008). | UC | North American[l] |
| *IL26* | IL-26 is secreted by Th17 cells, which mediate host defense against infections as well as tissue inflammation in chronic immune mediated diseases, including IBD (Aujla et al. 2007) | UC | North American[l] |
| *IL23R* | Key differentiation feature of $CD4^+$ Th17 cells, which areefector cells that are critical in mediating antimicrobial defenses (Abraham Clara et al. 2009) | CD, UC | North American[c,d,l], Belgian/French[e], United Kingdom[f,g,k], Quebec[i], German[j], |
| *MST1* | Macrophage stimulating 1 (MST1) encodes a protein inducing phagozytosis and involved in suppressing cell-mediated immunity by down-regulating IL-12 (Morrison et al. 2004). | CD, UC | United Kingdom[f,g,k] |
| *TNFRSF6B* | Involved in both antigen-presenting cell differentiation and lymphocyte function (Hsu TL et l. 2002; Chang et al. 2008). | IBD | North American (European: pediatric-onset)[m] |
| *PSMG1* | The proteasome assembly chaperone 1 (PSMG1) is involved in the maturation of the mammalian 20S proteasomes (Hirano et al. 2005). Modest increased in PSMG1 colonic expression was observed in IBD patients compared to controls (Kugathasan et al. 2008). | IBD | North American (European: pediatric-onset)[m] |

[a] Yamazaki K et al. 2005; [b] Hampe et al. 2007; [c] Duerr et al. 2006; [d] Rioux et al. 2007; [e] Libioulle et al. 2007; [f] Wellcome Trust Case Control Consortium 2007a; [g] Parkes et al. 2007; [h] Franke et al. 2007; [i] Raelson et al. 2007; [j] Franke et al. 2008a; [k] Fisher et al. 2008; [l] Silverberg et al. 2009; [m] Kugathasan et al. 2008; [n] Disease where associations were observed; [o] Studies reporting associations for each gene.

Although the effect size of each gene and locus identified so far, with the exception of *NOD2* and *IL23R*, is relatively modest (odds rations of 1.2-1.7 per risk allele) (Barrett et al. 2008), the importance of their discovery is highlighted by the newly uncovered genes, the pathways in which they are involved, and the discovery of previously unsuspected mechanisms that suggest a predisposition to chronic intestinal inflammation. In addition to the several new susceptibility loci identified by these GWAS, more importantly are the novel loci that were reported by more than one study as well as the distinct signals from different genes all part of or converging toward the same pathway. Indeed, some clear themes emerged, building on both the prior knowledge of association with *NOD2* and the recognition of the importance of bacterial drivers in contributing to intestinal inflammation. In addition to the importance of the role of *NOD2* in intracellular bacterial sensing to initiate the immune response, two novel key pathways also contributing to IBD pathophysiology have been uncovered and confirmed by several studies: the activation of naïve $CD4^+$ T cells by the IL-12/IL-23 pathways and the cellular process of autophagy (Cho 2008). Both pathways will be briefly reviewed.

#### 1.4.6.1 IL-23R and CD association

Amongst the first GWAS published for any disease, Duerr *et al.* reported a strong association between variants in the IL23 receptor (*IL23R*) region, located on chromosome 1p31, and CD and UC susceptibility (Duerr et al. 2006). This association result was subsequently replicated in several GWAS by groups from

Belgium-France (Libioulle et al. 2007), United Kingdom (Wellcome Trust Case Control Consortium 2007a; Parkes et al. 2007) and Quebec (Raelson et al. 2007) (see **Tables 1.1-1.2**). A rare, highly conserved amino acid polymorphism (Arg381Gln) located in the cytoplasmic domain of the receptor, was found to confer approximately three-fold protection against developing CD in heterozygous individuals, with a more modest protective effect observed in UC (Duerr et al. 2006). Multiple other *IL23R* variants, encompassing 7 of the 12 exons of *IL23R* as well as the intergenic region between *IL23R* and its homologue *IL12RB2*, showed additional independent associations (Duerr et al. 2006). The patterns of LD amongst these latter variants suggested the presence of multiple independent susceptibility alleles (Duerr et al. 2006). The *IL23R* associations were not replicated in a well-powered CD cohort from Japan (Yamazaki et al. 2007).

### *IL-23 signaling pathway and association with CD & UC*

The IL-23R complex is composed of the IL-23R and IL-12RB1 (chromosome 19p13) subunits, the latter also being part of the IL-12 receptor complex (Parham et al. 2002). The IL-23 cytokine is composed of p19 (IL23A, chromosome 12q13) and p40 (IL12B, chromosome 5q33) subunits, the latter being also part of the IL-12 cytokine (Oppmann et al. 2000). Following the engagement of IL-23R by IL-23, Janus kinase 2 (JAK2) is activated, resulting in JAK2 autophosphorylation and phosphorylation of IL-23R (Abraham et al. 2009). This in turn leads in the recruitment, phosphorylation, homodimerization, and nuclear translocation of STAT3 transcription factor, as well as the activation of

STAT1, STAT4, and STAT5 (Parham et al. 2002; Abraham et al. 2009). Interestingly, in a recently published meta-analysis of three large GWAS (Barrett et al. 2008), four regions among the top 32 associated with CD are known to play a direct role in IL-23 signaling, highlighting the importance of this pathway in the CD immunopathogenesis. These include *IL23R*, p40 (*IL12B*), *STAT3*, and *JAK2*, which were ranked respectively 1, 11, 12, and 30 in terms of significance (Barrett et al. 2008). Although the causal alleles have not been identified in these candidates because of the LD patterns in these regions, this meta-analysis has very likely identified the genes responsible for the association signals in these regions. Noteworthy, disease-associations with variants in the *IL23R* region have also been reported to confer susceptibility to UC (Duerr et al. 2006; Franke et al. 2008a; Franke et al. 2008b; Fisher et al. 2008; Silverberg et al. 2009), psoriasis (Cargill et al. 2007), and ankylosing spondylitis (Wellcome Trust Case Control Consortium et al. 2007b).

### *IL-23/IL-12 pathway: Role in determining naïve CD4$^+$ T cells differentiation*

Although no functional studies of the associated polymorphisms have yet been reported, these association results are indisputable and provide genetic corroboration of the recently published functional data highlighting the importance of this pathway in inflammation. Indeed, the IL-23/IL-12 pathway plays a key role in determining the differentiation of naïve CD4$^+$ T cells into effector T helper 1 (Th1) cells (driven by IL-12) or T helper 17 (Th17) cells

(driven by IL-23) (Abraham et al. 2009). The Th1 cells mediate immunity against intracellular microbes via the secretion of interferon gamma (INFγ) and TNF-α, resulting in macrophage activation and IgG production (Abraham et al. 2009). The Th17 cells are characterized by the production of IL-17, TNF-α, IL-6, IL-21, IL-22, and IL-26 (Murphy et al. 2003; Hue et al. 2006). These cells mediate defense against a broad range of pathogens, including extracellular Gram-positive and Gram-negative bacteria, as well as fungi (Bettelli et al. 2008). Th17 cells have also been shown to play a key role in driving organ-specific autoimmune inflammation in a number of animal models (Wiekowski et al. 2001).

### *Role of IL-23 in inflammation*

The association between *IL23R* and CD further supports the recently proposed role of IL-23 in mucosal inflammation, where IL-23 activation of the Th17 T cells results in production of IL-17 and promotion of chronic inflammation (Neurath 2007; Xavier et al. 2007). Noteworthy, the observation of increased levels of expression of IL-17 and IL-23 in the intestinal lamina propria of CD patients brings support to the *in vivo* importance of the IL-23 signaling pathway in human IBD (Fujino et al. 2003; Schmidt et al. 2005; Hölttä et al. 2008; Fina et al. 2008). Additionally, other murine models of colitis have further demonstrated the role of the IL-23 pathway in mediating intestinal inflammation either through deficiency or blockade of IL-23 (Kullberg et al. 2006; Yen et al. 2006; Hue et al. 2006; Uhlig et al. 2006; Elson et al. 2007; Izcue et al. 2008). For example, treating IL-10-deficient mice with monoclonal antibodies specific for p19 improved the spontaneous intestinal inflammation usually observed in these

mice (Wiekowski et al. 2001). This treatment was also shown to ameliorate the inflammation in the offspring of IL-10-deficient mice and p19-deficent mice crosses.

### *IL-23/IL-12: therapeutical targets*

It has been shown that blocking antibodies specific for p40 inhibit both IL-23 and IL-12 induced signaling, whereas p19-specific antibodies only inhibit IL-23 induced signaling (Abraham et al. 2009). More precisely, the administration of p40-specific antibodies has proven to be a promising approach for treating CD by blocking both the IL-23 and IL-12 activities (Mannon et al. 2004). However, since these pathways also play a key role in mediating anti-microbial defense mechanisms, the risk of infectious complication will need to be considered. One proposed solution is to specifically block the IL-23 pathway using a monoclonal antibody against p19 (Abraham et al. 2009). Such an antibody has proven to be effective in both the prevention and treatment of a murine model of $CD4^+$ T cell-mediated colitis (Elson et al. 2007), suggesting a potentially new effective therapeutic approach for treating CD patients (Elson et al. 2007).

## 1.4.6.2 CD and the authophagy pathway

The latest GWAS breakthroughs have expanded the role of the innate immunity components beyond the already implicated *NOD2* to include two separate members of the autophagy cellular process, *ATG16L1* and *IRGM* (Hampe et al. 2007; Parkes et al. 2007). Both these genes may play a key role in

the autophagy process by which cells recycle redundant organelles (Hampe et al. 2007; Parkes et al. 2007). More importantly for CD, this process also plays an important role in defense against intracellular micro-organisms by encapsulating/eliminating microbes that invade the cytosol (Birmingham et al. 2006; Nakagawa et al. 2004). Whether bacteria penetrate the cytoplasm by invasion or escape from endocytic vesicles; they are usually engulfed by the autophagy process (Deretic 2006). Once bacteria are compartmentalized within autophagosomes, the cell can use a variety of mechanisms to eliminate them, including the generation of a locally hostile environment within the vesicle by fusing with lysosomes and activating respiratory burst cascades (Radtke et al. 2006).

The autophagy process is an important part of the innate response to remove intracellular bacteria and viruses, and is also linked to adaptive immunity by degradation of intracellular proteins with subsequent MHC class II antigen processing (Schmid et al. 2006; Münz 2006). Interestingly, it has long been hypothesized that commensal intestinal bacterial predispose individuals to CD but without any good understanding as to how; defects in the autophagy process may be one of these missing mechanisms.

### *ATG16L1*

*ATG16L1* (autophagy-related 16-like 1), located on chromosome 2q37.1, is a member of the family of genes involved in the autophagy process. This gene is expressed in the intestinal epithelial cells, antigen-presenting cells, $CD4^+$,

$CD8^+$, and $CD19^+$ primary human T cells (Hampe et al. 2007; Rioux et al. 2007). Functional knockouts of this gene have been reported to markedly impair the autophagy process of the intracellular pathogen *S. typhimurium* in intestinal epithelial cells, and have demonstrated that *ATG16L1* is essential for the formation of autophagosomes in response to serum starvation or bacterial infection (Rioux et al. 2007).

The association between *ATG16L1* and CD was first identified (Hampe et al. 2007) using a 20k non-synonymous SNP array. This finding has since been replicated in at least two independent groups (**Tables 1.1-1.2**; Rioux et al. 2007; Parkes et al. 2007; Wellcome Trust Case Control Consortium 2007a). One variant (Ala281Thr/T300A/rs2241880) was found to be highly associated with CD, with the less common threonine allele conferring protection to CD. Following haplotype and regression analyses, this variant was found to explain the entire association signal observed at the *ATG16L1* locus. The exact functional impact of this variant remains unclear. The ATG16L1 protein comprises an N-terminal APG16 domain consisting of coiled-coil domains and eight C-terminal WD repeats (Hampe et al 2007). The Ala281Thr substitution is located at the N-terminus of the WD-repeat domain, which is an evolutionarily conserved domain (Hampe et al. 2007). A study showed that in cultured epithelial cells, the *ATG16L1*300A* allele had no effect on housekeeping autophagy but caused a defective autophagic response to invasive bacteria (Kuballa et al. 2008). In addition, the same group reported that the *ATG16L1*300A* protein was unstable in infected epithelial cells, a feature not seen with the protective *ATG16L1*300T* allele (Kuballa et al. 2008).

### *IRGM*

*IRGM* (immunity-related guanosine triphosphatase) encodes another important component of the autophagy process. The association between *IRGM* variants and CD susceptibility was first observed in the Welcome Trust Case Control Consortium (WTCCC) GWA scan (Wellcome Trust Case Control Consortium 2007a; Parkes et al. 2007). Sequencing of *IRGM* in CD patients and healthy controls did not identify any causal amino acid changes that could explain the association signal (Parkes et al. 2007; McCarroll et al. 2008). However, a recent publication by McCarroll *et al.* identified a common 20-kb deletion polymorphism located immediately upstream of *IRGM* and in perfect LD with the most strongly CD-associated SNP (McCarroll et al. 2008). They reported that the deletion (CD risk) and the reference (CD protective) haplotypes of *IRGM* showed distinct expression patterns and that manipulation of *IRGM* expression levels modulated the cellular autophagy of internalized bacteria (McCaroll et al. 2008). This group suggested that the CD associations at *IRGM* locus arose from an alteration in *IRGM* regulation, affecting the efficacy of autophagy (McCaroll et al. 2008).

*IRGM* belongs to an emerging family of genes encoding interferon-inducible guanosine triphosphatases (IRGs) involved in pathogen clearance (Taylor 2007). The protein encoded by *IRGM* exists in a number of isoforms and binds GTP to induce autophagy and subsequently generate large autolysosomal organelles as a mechanism for eliminating intracellular bacteria (Singh et al. 2006). Mice deficient in *IRGM* have an impaired capacity to eliminate the

intracellular pathogens *Toxoplasma gondii* and *L. monocytogenes* (Collazo et al. 2001). Moreover, *IRGM* knockdowns in human macrophages leads to markedly prolonged survival of *Mycobacterium tuberculosis*, demonstrating that *IRGM* is necessary to induce autophagy in order to efficiently eliminate intracellular mycobacteria (Singh et al. 2006).

## 1.5 IBD and genetics of the innate immune system

In general, the main genetic associations in IBD can be divided into genes that contribute to the innate immune response and those that contribute to the adaptive immune response. The trigger of the innate immunity seems to be a prerequisite for the excessive activation of the adaptive immune response, while the latter is thought to be the proximate driver of tissue damage that is observed in IBD patients (Xavier et al. 2007).

Confirmed genes of the innate immune subgroup comprise *NOD2* and the two autophagy-related genes, *ATG16L1* and *IRGM*. Interestingly, these three associations have been reported to be specific to CD as several studies showed no associations with UC, suggesting that both alterations in the recognition and intracellular processing of bacterial components constitute a central feature of CD pathogenesis (Cho et al. 2008; Fisher et al. 2008; Franke et al. 2008b). At the time the two first projects described in this thesis were initiated (Chapters 2-4), the knowledge in the field of IBD genetics was limited to the pre-GWAS era, and much focus was still on *NOD2* and its central role as an intracellular bacteria

sensor. Hence, the topics of the two first projects, focusing on genes encoding proteins also involved in intracellular bacterial sensing, were selected with the knowledge that the processes of monitoring the intestinal luminal contents and microbial flora as well as titrating the mucosal response were probably central to the functional integrity of the mucosa, and thus could be involved in the susceptibility of CD and UC.

## 1.6 The first family of PRMs: Toll-like receptor (TLR) family

The fact that commensal bacteria may influence the development and function of the immune system, coupled with evidence that alterations in the acquisition or composition of the microbial community can influence disease susceptibility, supports a model in which members of the mammalian innate immune system are constantly sampling the dynamic composition of commensal communities (Artis 2008).

For almost two decades, Charles Janeway and his colleagues have suggested that the innate immune system recognizes the signature of pathogens collectively called pathogen-associated molecular patterns (PAMPs) (Medzhitov et al. 1998; Janeway et al. 2002). In the past decade, the discovery of many microbial sensors termed "pattern-recognition molecules" (PRMs), also known as "pattern-recognition receptors" (PRRs), that are able to detect conserved microbial motifs known as PAMPs and to discern pathogens from commensals

has accelerated the research in the field of innate immunity (Medzhitov et al. 1998). Three major families of PRMs have been identified: the Toll-like receptors (TLRs), the RIG-I-like receptors (RLRs) that include the cytosolic helicases RIG-1 and MDA5 sensing viruses, and the Nod-like receptors (NLRs) (Meylan et al. 2006). The first PRM family of proteins identified was the TLRs (Pasare et al. 2005; Kawai et al. 2007), which are the homologs of the Toll molecule originally identified in *Drosophila melanogaster* (Anderson et al. 1985; Medzhitov et al. 1997). These transmembrane receptors are composed of an extracellular or luminal binding domain composed of leucine-rich repeats (LRRs) that recognize PAMPs and of a signal transduction domain, known as Toll/IL-1 receptor (TIR), pointing towards the cytoplasm (Akira 2001). The TIR domain is required for interaction with TIR-containing adaptor molecules, such as MyD88, MyD88-adaptor-like (MAL (also known as TIR domain-containing adaptor protein-TIRAP), TIR-domain-containing adaptor protein inducing INF$\beta$ (TRIF), and TRIF-related adaptor molecule (TRAM) (Pasare et al. 2005; Kawai et al. 2007). This latter interaction results in the activation of NF-$\kappa$B and activator protein 1 (AP-1) in the MyD88-dependent pathways and/or type I interferons (IFNs) in TRIF-dependent antiviral pathways (Creagh et al. 2006). Thirteen TLRs (named TLR1-TLR13) have been identified in humans and mice together, and equivalent forms have been identified in other mammalian species (Kawai et al. 2007).

## 1.7 The NOD-like receptor (NLR) family

It quickly became apparent that TLRs could not account for all features of the innate immune response as these receptors were unable to sense intracellular pathogens. The confirmation of the existence of an intracellular mode of bacterial detection came with the report that only an invasive form of the enteric bacterium *Shigella flexneri* activated the NF-κB pathway in epithelial cells (Philpott et al. 2000). Subsequent studies identified the cytosolic protein NOD1 (nucleotide-binding oligomerization domain 1) as the microbial sensor responsible for the epithelial cell's NF-κB-dependent response to *S. flexneri in vitro* (Girardin et al. 2001).

NOD1 is the founding member of the Nod-like receptors (NLRs) family, and in the past few years, bioinformatics analyses have identified the existence of over 22 NLR genes in the human genome (Harton et al. 2002; Werts et al. 2006; Ting et al. 2006). Indeed, since these receptors lack reported or predicted signal peptides and transmembrane domain in their amino acid sequences, NLRs are considered the cytoplasmic counterparts of TLRs. They are thought to localize in the cytosol where they can screen this compartment for PAMPs that have actively invaded the epithelial-cell barrier and provide a complementary surveillance system to the TLRs, which screen the luminal content (Fritz et al. 2006). Signaling synergy and crosstalk between the two PRM families has also been reported (Fritz et al. 2006).

Broadly, the NLR family comprises 5 members of the NOD (nucleotide-binding oligomerization domain) subfamily, 14 NALPs (NACHT domain, LRR

domain, and pyrin domain-containing protein), IPAF (IL-1β-converting enzyme [ICE]-protease-activating factor), NAIPs (neuronal apoptosis inhibitor factors), and CIITA (major histocompatibility complex [MHC] class II transactivator) (Ting et al. 2006). This family is also referred to as the CATERPILLER (CARD (caspase-recruitment domain) transcription enhancer, R (purine)-binding, pyrin, lots of leucine repeats) gene family, which comprises proteins with a NOD and a leucine-rich region (LRR) (Harton et al. 2002; Ting et al. 2006). Like the TLRs, the members of the NLR family of PRMs sense different bacterial components and toxins and, additionally, some members like NALP3 also sense endogenous danger signals or stress signals (danger-associated molecular patterns; DAMPs) released by dying or injured cells (Fritz et al. 2006; Ting et al. 2006).

### 1.7.1 The NLR family: domains

The members of the NLR family all share a characteristic tripartite structure consisting of 1) a C-terminus LRR domain that is possibly involved in the recognition of conserved microbial patterns or other ligands; 2) a centrally located NACHT domain [NACHT stands for domain present in Naip, CIITA, HET-E (plant het product involved in vegetative incompatibility) and TP-1 (telomerase-associated protein 1)] that mediates self-oligomerization and is essential for activation of the NLRs; and 3) a N-terminal effector domain, which is responsible for homophilic interactions with other proteins containing domains of the same class, resulting in signal transduction (Carneiro et al. 2008; Franchi et al. 2009). Based on the nature of the N-terminal domain, the NLRs have been

further divided in subfamilies: the NODs (NOD1 and NOD2) and IPAF possess a CARD; the NALPs (NALP 1-14) have a pyrin domain (PYD); and NAIP contains a baculovirus inhibitor of apoptosis protein repeat domain (BIR) (Werts et al. 2006; Carneiro et al. 2008; Franchi et al. 2009). A summary of the NLR family members classified by their domain type can be found in **Table 1.3**.

**Table 1.3: Summary table of the Nucleotide oligomerization domain-like receptor (NLR) family members**

| Nomenclature | Other names | Domain structure | Chromosomal location | GeneBank accession | Microbial motifs recognized |
|---|---|---|---|---|---|
| ***CARD domain*** | | | | | |
| ***NOD1*** | NLRC1;CARD4; CLR7.1 | CARD-NACHT-LRR | 7p14.3 | AF126484 | GM-tripeptide; c-D-Glu-DAP(iEDAP); D-lactyl-L-Ala-c-Glu-meso-DAP-Gly (FK156); heptanolyl-c-Glu-meso-DAP-Ala (FK565) |
| ***NOD2*** | NLRC2;CARD15; CLR16.3 | CARD-CARD-NACHT-LRR | 16q12.1 | AF178930 | MDP; MurNAc-L-Ala-g-D-Glu-L-Lys (M-TRILys) |
| ***CIITA type1*** | NLRA; MHC2TA;C2TA | CARD-(X-NACHT-LRR) | 16p13.13 | AF000002 | |
| ***PYRIN domain*** | | | | | |
| ***NLRP1*** | NALP1;CARD7;DEF CAP;NAC;CLR17.1 | PYD-NACHT-LRR-FIIND-CARD | 17p13.2 | AB023143 | MDP; Lethal Toxin ? |
| ***NLRP2*** | NALP2;PYPAF2;NB S1;PAN1;CLR19.9 | PYD-NACHT-LRR | 19q13.42 | AK000517 | |
| ***NLRP3*** | NALP3;PYPAF1; CIAS1;CLR1.1 | PYD-NACHT-LRR | 1q44 | AF054176 | Bacterial RNA; viral RNA;uric acid crystals; LPS;LTA;MDP |
| ***NLRP4*** | NALP4;PYPAF4;PA N2;RNH2;CLR19.5 | PYD-NACHT-LRR | 19q13.43 | AF479747 | |
| ***NLRP5*** | NALP5;NOD14;PYP AF8;MATER;PAN11; CLR19.8 | PYD-NACHT-LRR | 19q13.43 | AY154460 | |
| ***NLRP6*** | NALP6;PYPAF5;PA NS; CLR11.4 | PYD-NACHT-LRR | 11p15.5 | AF479748 | |
| ***NLRP7*** | NALP7;NOD12;PYP AF3;PAN7;CLR19.4 | PYD-NACHT-LRR | 19q13.42 | AF464765 | |
| ***NLRP8*** | NALP8;NOD16;PAN 4;CLR19.2 | PYD-NACHT-LRR | 19q13.43 | AY154463 | |
| ***NLRP9*** | NALP9;NOD6;PAN1 2;CLR19.1 | PYD-NACHT-LRR | 19q13.43 | AY154464 | |
| ***NLRP10*** | NALP10;NOD8;PYN OD;PAN5;CLR11.1 | PYD-NACHT | 11p15.4 | AY154465 | |
| ***NLRP11*** | NALP11;NOD17;PY PAF6;PAN10; CLR19.6 | PYD-NACHT-LRR | 19q13.42 | AY095145 | |
| ***NLRP12*** | NALP12;PYPAF7;M onarch1;RNO2; PAN6;CLR19.3 | PYD-NACHT-LRR | 19q13.42 | AY095146 | |
| ***NLRP13*** | NALP13;NOD14; PAN13;CLR19.7 | PYD-NACHT-LRR | 19q13.42 | AY154468 | |
| ***NLRP14*** | NALP14;NOD5; PAN8;CLR11.2 | PYD-NACHT-LRR | 11p15.4 | BK001107 | |
| ***BIR domain*** | | | | | |
| ***NAIP*** | NLRB1;BIRC1; CLR5.1 | BIR-BIR-BIR-NACHT-LRR | 5q13.1 | U19251 | |
| ***Untypical CARD*** | | | | | |
| ***NLRC4*** | CARD12;CLAN; IPAF;CLR2.1 | Card?-NACHT-LRR | 2p22-p21 | AF376061 | Flagellin from Salmonella, Legionella, Listeria, Pseudomonas |
| ***NLRC5*** | NOD27;CLR16.1 | Card?-NACHT-LRR | 16q13 | AF389420 | |
| ***NLRC3*** | NOD3; CLR16.2 | Card?-NACHT-LRR | 16p13.3 | BK001112 | |
| ***Undefined*** | | | | | |
| ***NLRX1*** | NOD9;CLR11.3 | X-NACHT-LRR | 11q23.3 | AB094095 | |

A similar domain organization is found in a large family of proteins in plants called R, in which the so-called "NBS-LRR" (nucleotide binding site-LRR) proteins are involved in the defense against plant pathogens (Belkhadir et al. 2004; DeYoung et al. 2006). The R protein recognizes specific products of microbial pathogens and elicits a plant-host response (Ausubel 2005).

### 1.7.2 The NLR family: mechanism of action

To date, only a few of these 22 NLR members have known ligands and, similar to the TLRs, the mechanism involved in PAMPs recognition by NLRs has not been characterized; no direct interaction has been demonstrated yet. It is therefore possible that the recognition is indirect and involves other proteins (Carneiro et al. 2008; Franchi et al. 2009). The biology of most mammalian NLRs is still ill-defined, but some NLRs like NOD2 (discussed previously) and NALP3 (discussed bellow) are better characterized.

The current models for NLR signaling suggest that NLRs are normally present in the cytoplasm in an inactive self-repressed form, where the LRR domain folds back onto the NACHT domain, thereby inhibiting spontaneous oligomerization and activation of the NLR protein (Carneiro et al. 2008; Franchi et al. 2009). Upon direct or indirect binding of the ligand to the LRR domain, the receptor undergoes a conformational rearrangement exposing the NACHT domain and thereby allowing oligomerization and subsequent exposure of the effector N-terminal domain (Martinon et al 2005; Carneiro et al. 2008; Franchi et al. 2009).

The PYD and CARD domains of NALPs and NODs, respectively, normally link these proteins to downstream adaptor or effector proteins through homotypic interactions. Subsequent to its activation, NLRs mediate pro-inflammatory signaling mainly through two major pathways: activation of transcription factor NF-κB and the caspase-1 inflammasome (discussed below) (Martinon et al. 2005; Carneiro et al. 2008; Franchi et al. 2009).

In the context of an infection, the pathogens are likely to activate the immune response through several PRMs simultaneously (Carneiro et al. 2008; Franchi et al. 2009). Even though some PRMs might play more relevant roles than others depending on the nature of the pathogens, it has been proposed that co-stimulation of multiple PRMs is responsible for a full-blown microbial-induced immune response *in vivo* (Carneiro et al. 2008; Franchi et al. 2009). However, it is still not clear how all the signals elicited by distinct PRM activation are integrated in one global immune response.

### 1.7.3 The NLR family & inflammatory human diseases

The understanding of the *in vivo* function of NLRs has deepened with studies reporting that mutations in some human NLR family members were associated with susceptibility to inflammatory disorders that have a high degree of penetrance (Ting et al. 2006). These findings have further strengthened the idea that these molecules play central roles in the regulation of inflammation and immunity.

Hereditary periodic fever syndromes (HPFS) are a group of monogenic auto-inflammatory syndromes unified by a central phenotype of recurrent inflammation, often manifested as unexplained fever (Ryan et al. 2008a). Known HPFS include the familial Mediterranean Fever (FMF), the tumor necrosis factor (TNF) receptor-associated periodic syndrome (TRAPS), the hyperimmunoglobulinemia D with periodic fever syndrome (HIDS), and the cryopyrin-associated periodic syndromes (CAPS) described below (Masters et al. 2006a; Ryan et al. 2008a). Disease-based gene discovery for these rare conditions has led to the description of novel pathways involved in the innate immune response and provided new insights into the molecular and cellular mechanisms implicated in the regulation of infection, endogenous stress, and auto-inflammation.

In addition to being involved in intracellular bacterial sensing, the two genes on which the two first projects focus (Chapters 2-4), *NLRP3* and *MEFV*, may contain mutations that can lead to HPFS (Ryan et al. 2008a). We hypothesized that while severe mutations in these genes will cause the respective monogenic syndrome, milder genetic changes such as SNPs may give rise to more subtly altered proteins, and hence contribute to the susceptibility of chronic inflammatory diseases like CD and UC.

For example, DNA sequence variants in the NALP1 gene (called *NLRP1*) are associated with increased risk of vitiligo-associated multiple autoimmune disease (Jin et al. 2007). As discussed previously, *NOD2* gained notoriety from the two studies (Ogura et al. 2001; Hugot et al. 2001) reporting associations with susceptibility to CD. Mutations in *NOD2* are also associated with Blau syndrome,

a rare autosomal dominant genetic disorder not related directly to CD that is characterized by granulomatous arthritis, uveitis, and skin lesions (Miceli-Richard et al. 2001). Four distinct missense mutations in NOD2 have been associated with the onset of this syndrome: R334Q, R334W, L469F, and E383E (Miceli-Richard et al. 2001; van Duist et al. 2005). R334Q and R334W are the most prevalent mutations and are located in the region encoding the NACHT domain (Miceli-Richard et al. 2001; van Duist et al. 2005). These mutations have been reported to result in a "gain-of-function" phenotype characterized by increased basal MDP-independent NF-κB activity (Miceli-Richard et al. 2001; Carneiro et al. 2007). It has thus been suggested that the inflammation observed in Blau syndrome patients could be the consequence of aberrant activation of NOD2 in absence of pathogenic stimuli (Miceli-Richard et al. 2001). *NOD2* mutations have also been implicated in early-onset sarcoidosis, a disease with similar clinical features to Blau syndrome (Kanazawa et al. 2005).

Interestingly, some residues in the NOD2 NACHT domain associated with Blau syndrome correspond to the same positions of pathogenic mutations in the NACHT domain of the closely related NALP3 protein (encoded by the *NLRP3* gene) that are associated with susceptibility to three rare and severe auto-inflammatory disorders, namely Muckle-Wells syndrome (MWS), familial cold auto-inflammatory syndrome (FCAS) and neonatal-onset multisystem inflammatory disease (NOMID) (Agostini et al. 2004; van Duist et al. 2005; Mariathasan et al. 2007). For example, the R334W mutation in the NACHT domain of NOD2 found in Blau syndrome patients corresponds structurally to the R260W mutation located in the NACHT domain of NALP3 (Mariathasan et al.

2007). The R260W change in NALP3 has been shown to confer a "gain-of-function" phenotype, resulting in a constitutively active protein, over-activation of caspase-1, and aberrant maturation and release of IL-1$\beta$ from peripheral blood leukocytes of MWS patients (Agostini et al. 2004).

Given the importance of bacterial sensing in CD pathogenesis, the association of *NOD2* mutations with CD susceptibility, the role of NOD2 as a PRM, and the mutations within the NOD2 NACHT domain associated with Blau syndrome that are structurally associated with NALP3 mutations responsible for three auto-inflammatory disorders, we hypothesized that *NLRP3* could also be a good candidate gene participating to CD susceptibility (Chapters 2, 4).

## 1.8 *NLRP3* gene & NALP3 protein

*NLRP3* (nucleotide binding oligomerization domain, leucine-rich-repeat family, pyrin containing 3) (also known as *CIAS1* and *NALP3*), the gene encoding NALP3 (NACHT, leucine-rich repeat and pyrin domains containing protein 3) (also known as cryopyrin), was first identified in 2001 by positional cloning in the search for the genetic basis of two autosomal-dominant auto-inflammatory diseases: FCAS and MWS (Hoffman et al. 2001a). The identification of a linkage peak located in the 1q44 chromosomal region with no previously recognized genes was made possible by the availability of large multi-generational families from North America and Europe (Hoffman et al. 2001a). Subsequent screening of expressed sequence tags (ESTs) in this region has allowed for the discovery of

heterozygous missense mutations in the *NLRP3* gene in patients with these two similar disorders (i.e. FCAS and MWS), which were previously considered as distinct (Hoffman et al. 2001a; Agana et al. 2002; Dodé et al. 2002). Similarities in the clinical phenotypes of MWS and NOMID prompted the discovery of *NLRP3* mutations in these patients as well (Aksentijevich et al. 2002; Feldmann et al. 2002).

### 1.8.1 Clinical manifestations of MWS, FCAS, and NOMID

These three disorders have been identified in patients around the world. FCAS is widely spread in North America, which is explained by a founder effect (Hoffmann et al. 2003), and MWS seems to be more prevalent in Europe (Ting et al. 2006). These *NLRP3*-associated diseases are classified as auto-inflammatory disorders, which are a new disease classification that is characterized by recurrent episodes of systemic inflammation in the absence of infection, and are distinguished from traditional autoimmune diseases by the lack of production of high-titre auto-antibodies or antigen-specific T cells (Ting et al. 2006). FCAS, MWS and NOMID were originally described as distinct clinical entities despite the fact that their symptoms overlap (Hawkins et al. 2004; Neven et al. 2004; Granel et al. 2005; Hentgen et al. 2005); patients often present with fever, pseudo-urticarial skin rash, and joint involvement of varying severity associated with neutrophil-mediated inflammation and an intense acute-phase response (Neven et al. 2008). These three diseases are now considered to be a continuum of one disease that is referred to as cryopyrin-associated periodic syndromes (CAPS),

with FCAS being the mildest, MWS being of intermediate severity, and NOMID being the most severe (Neven et al. 2008). Although these three disorders share some clinical characteristics, they also have distinguishing features, which will be briefly reviewed.

#### 1.8.1.1 Familial cold auto-inflammatory syndrome (FCAS)

FCAS, also known as familial cold urticaria (FCU), was first described by Kile and Rusk in 1940 (Kile et al. 1940). This autosomal dominant syndrome is characterized by recurrent, short, self-limited episodes of low-grade fever, rash and arthralgia that are precipitated upon exposure to cold temperature (Hoffman et al. 2001b; Hofmann et al. 2003; Johnstone et al. 2003). The symptoms usually begin 1-2h after generalized exposure to cold temperature, and the duration of the attacks is usually short, lasting less than 24 hours (Neven et al. 2008).

#### 1.8.1.2 Muckle-Wells syndrome (MWS)

Muckle and Wells first described this syndrome in 1962, as they reported members of one family who presented with urticaria, deafness and renal amyloidosis (Muckle et al. 1962). The disease is characterized by recurrent episodes of fever and rash associated with joint and eye manifestations (Neven 2008). Precipitation factors can usually not be identified and triggering by cold temperature is rarely observed (Neven 2008). The perception of deafness is common, occurring in approximately 70% of cases, and it usually begins in childhood or early adulthood (Neven 2008). Amyloidosis caused by chronic

inflammation is the most serious complication of MWS and may develop in adulthood in 25% of the cases. (Neven 2008).

#### 1.8.1.3 Neonatal-onset multisystem inflammatory disease (NOMID)

NOMID (Neonatal-onset multisystemic inflammatory disease), also referred to as CINCA (chronic infantile neurological cutaneous articular syndrome), was first described by Prieur in the early 1980s as a chronic inflammatory disease with rash, articular involvement, and chronic aseptic meningitis (Prieur et al. 1980; Prieur et al. 1981). The first symptoms usually occur at birth or early infancy. Abnormalities of the central nervous system are present in almost all patients and are caused by chronic, aseptic meningitis in which polymorphonuclear cells infiltrate the cerebrospinal fluid (Neven 2008). Severe disabilities are frequent, and premature death is possible in severely affected patients (Neven 2008).

### 1.8.2 Genetic risk factors for FCAS, MWS, and NOMID

FCAS, MWS, and NOMID are all caused by dominantly inherited or *de novo* mutations in *NLRP3*. Approximately 60 disease-associated mutations have been reported and most of them are clustered in the region encoding the NACHT domain and its flanking structure (corresponding to exon 3), suggesting that these mutations might alter the function of this domain (Milhavet et al. 2008). Since this domain is thought to mediate the oligomerization of NALP3, it has been hypothesized that these associated mutations might result in spontaneous NALP3

oligomerization (McDermott et al. 2007). This latter hypothesis is further supported by the facts that the pathogenic CAPS mutations are mainly "gain-of-function", resulting in aberrantly high production of IL-1$\beta$ (Hoffman et al. 2001a; Aganna et al. 2002; Dodé et al. 2002; Martinon et al. 2007), and that monocytes from MWS patients have been reported to secrete more mature IL-1$\beta$ than healthy donors even in the absence of NALP3 agonists (Agostini et al. 2004).

Approximately 40% of patients with NOMID do not carry any known *NLRP3* mutations, suggesting that the CAPS are genetically heterogenous (Martinon et al. 2007).

### 1.8.3 Disease management

The advance in the understanding of the genetic basis and disease mechanisms of these three disorders has lead to the development of new therapeutic strategies that target the interleukin-1 (IL-1) pathway. Anakira treatment has led to the most striking and dramatic improvements in the symptoms of all three disorders (Hawkins et al. 2003; Hoffman et al. 2004; Goldbach-Mansky et al. 2006). It is a recombinant, nonglycosylated homolog of the human IL-1 receptor antagonist that competitively inhibits binding of IL-1$\alpha$ and IL-1$\beta$ to the IL-1 receptor (Church et al. 2008). The success of the IL-1R blockade strategy in these conditions further supports the role of IL-1$\beta$ in these pathologies (Dinarello 2005a).

### 1.8.4 IL-1β: pivotal pro-inflammatory cytokine

IL-1β belongs to a cytokine family that includes IL-1α, IL-18, IL-33, and IL-1 receptor antagonist (IL-1RA). IL-1β was first cloned in the early 1980s (Auron et al. 1984) and is now considered the prototypical "multi-functional cytokine" affecting nearly all cells types, either alone or in combination with other cytokines, and plays a key role in the initiation and /or amplification of several biological processes, including lymphocyte activation and proliferation, endothelial cell activation, innate immune response, defense against pathogens, and inflammation (Dinarello 2005b). IL-1β is produced as an inactive pro-molecule by immune cells such as macrophages, monocytes, and dentritic cells, which must be cleaved into active IL-1β and released out of the cell to perform its function (Black et al. 1989; Kostura et al. 1989). This cleavage is catalyzed by the enzyme caspase-1, a cysteine protease that was first identified in 1991 (Thornberry et al. 1991). The activated caspase-1 cleaves a 116-amino acid region from the N-terminus of the cytosolic pro-IL-1β (p31) to convert it to the active form, IL-1β (p17) (Dinarello 1998). The role of caspase-1 in the processing of IL-1β is further supported by the fact that mice lacking caspase-1 are unable to produce mature IL-1β, and are therefore less susceptible to endotoxin induced shock (Vasilakos et al. 1995; Li et al. 1995; Kuida et al. 1995). The synthesis and release of IL-1β is tightly regulated to prevent its excessive production or dysregulation (Dinarello 2007). Dysregulated release of IL-1β can be detrimental as this cytokine is considered to be one of the most potent endogenous pyrogens, causing fever, hypotension, and initiating host immune

defenses (Dinarello 1998). The understanding of the diverse biological activities in which IL-1$\beta$ is involved have led to a better understanding of the pathology of several diseases, including IBD (Ferrero-Miliani et al. 2007), gout (Martinon et al. 2006a), and many auto-inflammatory diseases including MWS, FCAS, NOMID, familial Mediterranean fever (FMF), pyogenic sterile arthritis with pyoderma gangrenosum and acne (PAPA) (Church et al. 2008).

### 1.8.5 NALP3 expression

NALP3 is expressed primarily in granulocytes, monocytes, dentritic cells, B and T cells, which are all important effectors cells of the immune system (Manij et al. 2002; Hoffman et al. 2001a; Krummer et al. 2007). NALP3 is also expressed in human chondrocytes (Krummer et al. 2007), and in the skin and eyes in mice (Feldmann et al. 2002; Anderson et al. 2004); these are the same tissues in which CAPS patients experience symptoms such as cartilage overgrowth, inflammatory rash and ocular inflammation (Neven 2008). However, the function of NALP3 in these tissues is still under study and has not been completely elucidated.

### 1.8.6 Discovery of the first inflammasome complex: regulation of caspase-1

Caspases are cysteine proteases that play a crucial role in regulating apoptosis and cytokine activation (Martinon et al. 2007b ; Yu et al. 2008). Three genes encode inflammatory caspases in humans: caspase-1, caspase-4, and

caspase-5 (Martinon et al. 2007b ; Yu et al. 2008). Caspase-1 itself is synthesized as a catalytically inactive 45-kDa zymogen (i.e. pro-caspase-1) that undergoes auto-catalytic processing following an appropriate stimulus. The active form of the enzyme, comprising the subunits p20 and p10, assembles into a heterotetramer, allowing the processing of pro-inflammatory cytokines such as IL-1$\beta$ (Martinon et al. 2007b; Yu et al. 2008). In 2002, Tshopp and colleagues, who were studying caspase-1 mechanism of activation (Martinon et al. 2002), based their experimental hypothesis on the fact that activation of related proteins caspase-8 and caspase-9 involved pro-caspase recruitment to oligomerized adaptors in a multiprotein complex (Tschopp et al. 1998; Chu et al. 2001). They investigated whether processing of caspase-1 also involved a multiple-adaptor complex. They reported the formation of a complex comprising the intracellular adaptor protein ASC (apoptosis-associated speck-like protein containing a CARD; also known as PYCARD, CARD5 or TMS1), the intracellular sensor NALP1, together with caspase-1 and caspase-5 in their *in vitro* experimental settings (Martinon et al. 2002). Since that first report, other inflammasome complexes have been identified and are defined by the NLR that they contain (Ogura et al. 2006; Drenth et al. 2006; Martinon et al. 2007b). The NALP3 inflammasome is one of them.

### 1.8.7 NALP3 inflammasome

The NALP3 inflammasome multiprotein cytoplasmic complex is very similar to the originally described NALP1 inflammasome. It also mediates the

activation of inflammatory caspase-1 and regulation of the IL-1$\beta$ and IL-18 post-translational processing (Martinon et al. 2002). Among its components are NALP3, the intracellular adaptor ASC, caspase-1, and CARDINAL (CARD-inhibitor of NF-κB activating ligand), which is thought to be a functional and structural homologue of the NALP1 C-terminus (Mariathasan et al. 2007). ASC encodes a 22-kDa protein that contains an amino (N)-terminal pyrin domain (PYD) and a carboxy (C)-terminal CARD (Masumoto et al. 1999; Conway et al. 2000). ASC plays a central role in the formation of the inflammasome complex, acting as an adaptor protein for the recruitment of other PYD- and CARD-containing proteins (Martinon et al. 2002). Through homotypic protein-protein interactions with its two domains, ASC can act as a bridge between the NALPs' sensor and the downstream caspase-1 effector (Martinon et al. 2007b).

The NALP3 polypeptide consists of three domains: a pyrin domain, a NACHT domain, and 11 LRR domains (Franchi et al. 2009). The pyrin and nucleotide binding domains of NALP3 are involved in protein-protein interactions in the inflammasome (Martinon et al. 2002). NALP3, which is probably the best understood member of the NALP family, is involved in the recognition of numerous exogenous and host ligands (i.e. PAMPs and DAMPs) that will be reviewed below (Martinon et al. 2007b).

In its inactive state, the LRR domain of NALP3 is thought to self-associate, preventing interaction with the adaptor proteins CARDINAL and ACS (Martinon et al. 2007b). Indeed, NALP3 has been reported to be initially associated with an HSP90-SGT1 complex that keeps the inflammasome inactive but competent for activation (Mayor et al. 2007). HSP90 binds to the NACHT

domain, affecting the stability of NALP3 by sustained inhibitions with HSP90 (Mayor et al. 2007). Once the cytoplasmic PAMP or DAMP is sensed by the LRR domain, HSP90 and SGT1 dissociate and the receptor undergoes a conformational rearrangement, exposing the NACHT domain and thereby allowing oligomerization that will result in the exposure of the effector PYD N-terminal domain (Martinon et al. 2005; Mayor et al. 2007). Once activated and unfolded, NALP3 enables the assembly of the inflammasome components CARDINAL, ASC and pro-caspase-1 through homotypic interactions between their respective PYD and CARD domains (Martinon et al. 2007; Martinon et al. 2007b; Church et al. 2008) (**Figure 1.3**). The PYD domain of NALP3 binds to the adaptor ASC via a PYD-PYD interaction and the CARD domain of ASC then recruits pro-caspase-1 to the inflammasome through a homotypic CARD-CARD interaction (Martinon et al. 2002; Srinivasula et al. 2002). A second pro-caspase-1 can be recruited via the C-terminal CARD of CARDINAL, forming a dimer with the pro-caspase-1 recruited by the ASC adapter (Agostini et al. 2004), while the FIIND domain (i.e. domain with function to find ) of CARDINAL interacts with the NAD domain (NACHT associated domain) of NALP3. The oligomerization of the inflammasome complex allows cross-activation and cleavage of pro-caspase 1 to its active form, caspase-1, which is then capable of cleaving the precursors pro-IL-1$\beta$ and pro-IL-18 to their active form, IL-1$\beta$ and IL-18 (Martinon et al. 2007; Martinon et al. 2007b) (**Figure 1.3**)

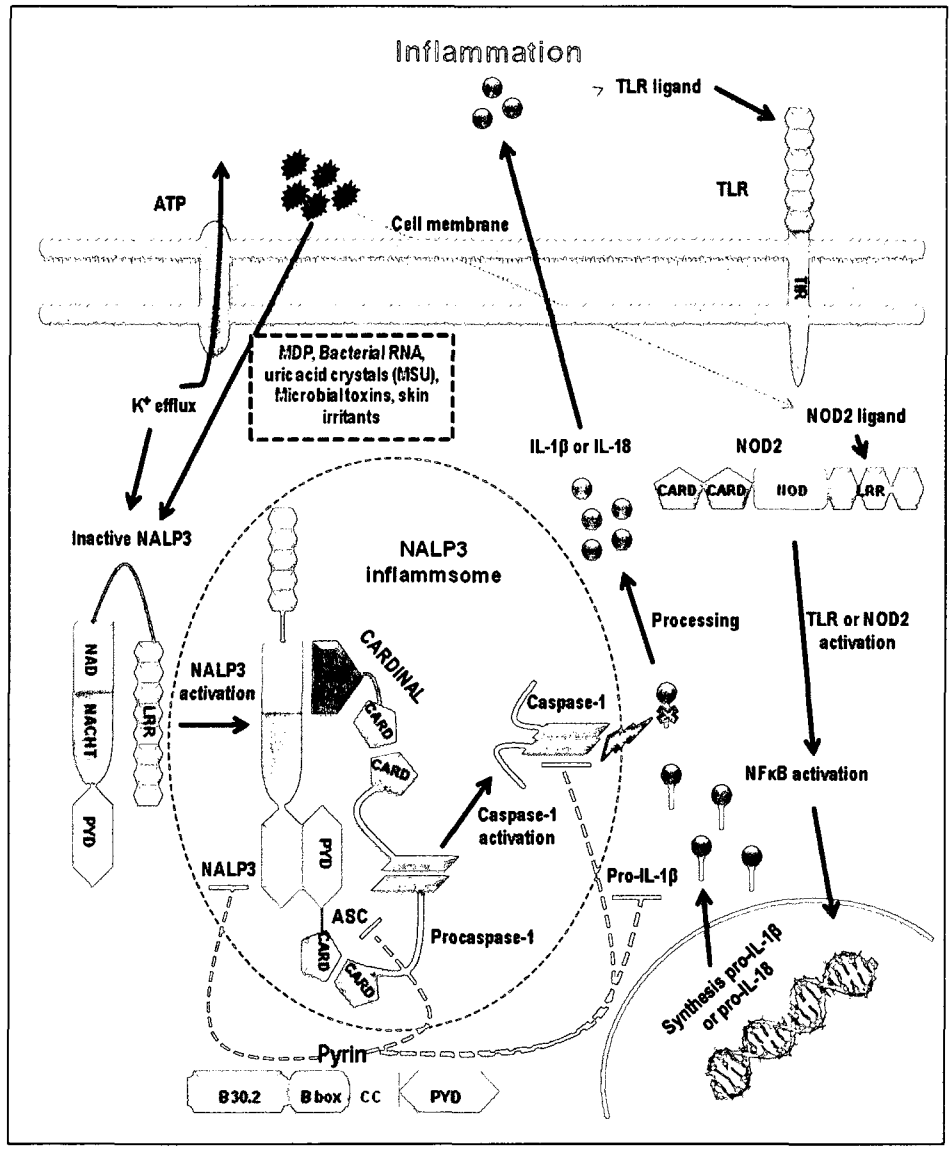



**Figure 1.3: Activation of the NALP3 inflammasome**

Engagement of specific TLRs and/or NOD2 leads to the activation of NF-κB signaling pathways, allowing the induction of pro-IL-1$\beta$ and/or pro-IL-18. Upon activation of NALP3 by agonists such as ATP, MDP, MSU, bacterial messenger RNA, or skin irritants, NALP3 molecule unfolds, enabling the assembly of the inflammasome components ASC, CARDINAL, and pro-caspase-1 through homotypic interactions between their pyrin (PYD) and CARD domains. The oligomerization of the inflammasome complex leads to cross-activation of pro-caspase-1 to its active form, caspase-1, which in turn

can cleave pro-IL-1$\beta$ to its active form, IL-1$\beta$. The pyrin protein has been reported to interact with NALP3, ASC, caspase-1, and pro-IL-1$\beta$. Abbreviations: ASC, apoptosis-associated speck-like protein containing a CARD; CARD, caspase-recruitment domain; CC, coiled-coil; FIIND, domain with function to find; IL, interleukin; LRR, leucine-rich repeat; MDP, muramyl dipeptide; MSU, monosodium urate; NACHT, domain present in neuronal apoptosis inhibitor protein (NAIP), major histocompatibility complex class II transactivator (CIITA), HET-E (incompatibility locus protein from *Podospora anserina*) and telomerase-associated protein (TP-1); NAD, NACHT-associated domain; NALP, NACHT domain, LRR domain, and pyrin domain-containing protein; NFκB, nuclear factor κB; NOD, nucleotide-binding oligomerization domain; NOD2, nucleotide-binding oligomerization domain containing 2; PYD, pyrin domain; TIR, Toll/IL-1 receptor; TLR, toll-like receptor. Figure adapted from (Church et al. 2008).

Knowledge of the components involved in the NALP3 inflammasome came from immunoprecipitation experiments, which demonstrated that NALP3 interacts with both ASC and CARDINAL, leading to the release of IL-1$\beta$ (Agostini et al. 2004). Moreover, NALP3 has been reported to co-localize with ASC in cytoplasmic specks that are visible by cytochemistry (Manij et al. 2002; Yu et al. 2006). *In vivo* studies with NALP3- and ASC-null mice, as well as *ex vivo* studies using macrophages from these mice, demonstrated that NALP3 and ASC are both necessary components for the activation of inflammasome-mediated inflammation (Kanneganti et al. 2006a; Mariathasan et al. 2006; Martinon et al. 2006a).

### 1.8.8 NALP3 inflammasome and IL-1β production

It is now generally accepted that activation and release of IL-1β requires two distinct signals (Church et al. 2008). What constitutes these signals *in vivo* during an infection or an auto-inflammatory response is still unclear. However, many studies demonstrated that the first signal can be provided by TLR activation, resulting in the NF-κB-dependent synthesis of pro-IL-1β (p35) (Martinon et al. 2007; Martinon et al. 2007b; Church et al. 2008). Examples of such activators leading to production of pro-IL-1β include LPS (TLR4 agonist), lipoteichoic acid (TLR2 agonist), CpG dinucleotides (TLR9 agonist), the antiviral compound R848 (also known as resiquimod; a TLR7 agonist), and the synthetic tripalmitoylated lipopeptide Pam3CysSerLys4 (Pam3CSK4; a TLR2 agonist) (Mariathasan et al. 2004; Dinarello 2005b; Mariathasan et al. 2006; Mariathasan et al. 2007; Franchi et al. 2009). A second signal is needed to activate caspase-1 for cleavage and secretion of active IL-1β (p17), which can be fulfilled by the NALP3 inflammasome complex (Church et al. 2008). Indeed, it has been demonstrated that IL-1β and IL-18 secretion is fully abrogated in macrophages from ASC-deficient mice in response to various TLR agonists (Mariathasan et al. 2004), and cytokine levels were reported to be markedly reduced in NALP3-deficient macrophages (Mariathasan et al. 2006). The requirement of a second signal for IL-1β production might constitute a "fail-safe" mechanism, ensuring that the activation of potent inflammatory responses occurs only in the presence of bona fide stimulus, such as the presence of invasive pathogens (not commensal bacteria alone) and/or tissue injury (Mariathasan et al. 2007). In auto-

inflammatory disorders, this tight regulation seems to be de-regulated, thereby leading to the inappropriate activation of IL-1$\beta$ in the absence of infection (Mariathasan et al. 2007).

### 1.8.9 NALP3 inflammasome activators: PAMPs

Bacteria are known to be potent activators of the NALP3 inflammasome and several bacterial PAMPs, such as MDP, LPS, and bacterial RNA have been reported to activate it (Pétrilli et al. 2007a). MDP, also an activator of NOD2, has been shown *in vivo* to induce IL-1$\beta$ release by NALP3, which was dependent on the presence of NOD2 (Pan et al. 2007). It is possible that activation of NOD2 by MDP then leads to the activation and/or recruitment of the NALP3 inflammasome, which in turn activates caspase-1 (Martinon et al. 2004). However, it is currently unclear whether NOD2 and NALP3 are two independent systems of MDP detection, or if NALP3 acts downstream of NOD2 in specific pathways leading to the inflammasome activation (Martinon et al. 2004). NOD2 and NALP3 may thus cooperate either directly or indirectly for IL-1$\beta$ activation and secretion (Pan et al. 2007).

Kannegati *et al.* provided evidence that LPS and other known TLR agonists, such as lipoteichoic acid (LTA) and Pam3CSK4, activate the NALP3 inflammasome (Kanneganti et al. 2006a). Yet, it was reported that some commercial sources of LPS are contaminated with PGN, and that PGN activates NALP3 and not pure LPS (Martinon et al. 2004). In answer to this controversial result, Kanneganti and colleagues subsequently demonstrated that direct

cytoplasmic delivery of LPS via the pore-forming streptolysin O is required for NALP3 activation (Kanneganti et al. 2007).

Using NALP3-deficient mice, Kannegati *et al.* also described the activation of the NALP3 inflammasome in response to bacterial RNA, such as those derived from *Escherichia coli*, *Listeria monocytogenes* and *Legionella pneumophila* (Kanneganti et al. 2006b), to small antiviral imidazoquinoline compounds R837 and R848, as well as to polyI:C, a synthetic double-stranded RNA (dsRNA) mimetic (Kanneganti et al. 2006a).

Additionally to these PAMPs, some microbes have been reported to be sensed by and activate the NALP3 inflammasome by virtue of their pore-forming toxins. These include Nigericin (*Streptomyces hygroscopicus*), aerolysin (*Aeromonas hydrophila*), maitotoxin (*Marina dinoflagellates*), gramicidin (*Bacillus brevis*), and $\alpha$-toxin (*S. aures*) (Walev et al. 1995; Mariathasan et al. 2006; Gurcel et al. 2006).

Moreover, two RNA viruses (*Sendai* and *Influenza*) have also been reported to activate the NALP3 inflammasome (Kanneganti et al. 2006b), but the viral PAMP responsible for the triggering of its activation is still unknown.

### 1.8.10 NALP3 inflammasome activators: DAMPs

Besides sensing the presence of PAMPs, NALP3 has the unique capacity to detect endogenous danger and stress signals referred to as DAMPs. In fact, the best-studied model of caspase-1 activation and subsequent IL-1$\beta$ release is from the exposure of cells to the DAMP ATP (adenosine triphosphate), and not to a

PAMP (Perregaux et al. 1994). In addition to ATP released by damaged cells, recent findings have suggested that NLRP3 is also responsible for activating caspase-1 via an inflammasome assembly in response to uric acid and to potassium efflux (Martinon et al. 2005; Martinon et al. 2006a; Martinon 2008).

Immunologists have long predicted that one of the major functions of the immune system is to differentiate self from non-self and to respond to self with tolerance while mounting an immune response against non-self components (Matzinger 1994; Matzinger 2002). Polly Matzinger and colleagues have proposed an alternative hypothesis, suggesting that the immune system may respond with tolerance to most antigens, and that the immune response would be triggered only upon the presentation of an antigen in the context of a danger signal, rather than being based on the foreignness of the antigen alone (Matzinger 1994; Matzinger 2002).

This latter hypothesis is important in the context of the intestinal microbial flora. The great majority of extracellular microbes living in the GI tract acts as commensals and do not constitute a threat to the host (Sonnenburg et al. 2004; Bäckhed et al. 2005). For the intestinal innate immune system not to be constantly activated, the microbes should be considered dangerous only if damage to the cell membrane, cell invasion, or metabolic changes in the local microenvironment is detected (Sonnenburg et al. 2004; Bäckhed et al. 2005). It is thus important for NALP3 to also detect non-microbial DAMPs molecules (Martinon et al. 2008), whose presence is often caused by invading pathogens.

#### 1.8.10.1 DAMPs: ATP

Extracellular ATP, which is often released from cells as a consequence of cell damage or non-apoptotic cell death, is another well-characterized danger signal that activates NALP3 and caspase-1 (Communi et al. 2000). ATP release may be triggered by different mechanical stimuli such as fluid shear stress (Milner et al. 1990; Grierson et al. 1995), stretch (Grygorczyk et al. 1997), hypotonic shock (Hazama et al. 1999), hydrostatic pressure changes (Ferguson et al. 1997), and compression (Sauer et al. 2000). Furthermore, the generation of ASC- (Mariathasan et al. 2004) and NALP3- (Mariathasan et al. 2006; Sutterwala et al. 2006) deficient mice demonstrated that ATP-mediated caspase-1 activation required both ASC and NALP3. However, fairly high ATP concentrations (2 to 5 mM) were required for these *in vitro* studies (Mariathasan et al. 2004; Mariathasan et al. 2006; Sutterwala et al. 2006). Taking into account that in an *in vivo* setting, most of the extracellular ATP could be rapidly hydrolyzed by ectonucleotidases (Eckle et al. 2007) further experiments need to be done to assess the physiological relevance of these findings.

For ATP-induced caspase-1 activation and subsequent IL-1$\beta$ release, several studies have also demonstrated the requirement of purinergic receptors of the $P2X_7$ subtype, for which ATP is thought to be the main endogenous ligand (Hogquist et al. 1991; Ferrari et al. 2006). The purinergic receptors are a family of plasma-membrane molecules that are involved in several cellular functions, such as vascular reactivity, apoptosis and cytokine secretion (Solle et al. 2001). The $P2X_7$ receptor activation usually results in potassium efflux, plasma-

membrane depolarization, cell swelling, and disaggregation of the cytoskeletal network (Ferrari et al. 2006). Likewise, it has been shown that depletion of intracellular potassium is crucial for the activation of the NALP3 inflammasome and caspase-1 (Perregaux et al. 1994; Walev et al. 1995; Pelegrin et al. 2006; Pelegrin et al. 2007; Kanneganti et al. 2007).

#### 1.8.10.2 DAMPs: MSU crystals, gout & pseudogout

Gout and pseudogout are two auto-inflammatory diseases that are characterized by arthropathies generated by the inflammatory reaction to the presence of microcrystals in the joints (Martinon et al. 2006b). More specifically, gout is caused by deposition of monosodium urate (MSU) crystals in joints and periarticular tissues, whereas pseudogout is caused by deposition of calcium pyrophosphate dihydrate (CPPD) crystals (Martinon et al. 2006b). In addition to its role in gout, MSU crystals have also been identified as a DAMP; these crystals are released from necrotic cells and are known to trigger IL-1$\beta$ secretion as well as innate immunity (Shi et al. 2003). More recently, Martinon and colleagues demonstrated that the NALP3 inflammasome and subsequent aberrant caspase-1 activation are linked to these two auto-inflammatory syndromes (Marinon et al. 2006). They reported that both MSU (in the case of gout) and CPPD (in the case of pseudogout) are detected by NALP3 as DAMPs, leading to the inflammasome assembly (Marinon et al. 2006). Additionally, macrophages from mice deficient in various components of the inflammasome, such as pro-caspase-1, ASC, or NALP3, did not respond to injection of MSU crystals, showing a reduced crystal-

induced IL-1$\beta$ secretion, highlighting the key role of the NALP3 inflammasome in these syndromes as well as the role of NALP3 as a DAMP sensor (Marinon et al. 2006).

#### 1.8.10.3 DAMPs: Skin irritants & UV irradiations

Skin irritants and ultraviolet irradiations are danger signals that can also trigger NALP3 inflammasome assembly. Indeed, ASC- and NALP3-deficient mice showed an impaired contact hypersensitivity response, which is a T cell-mediated immune response (Grabbe Set al. 1998), to the irritants trinitrophenylchloride (TNP-C) (Sutterwala et al. 2006), 2,4-6-trinitrochlorobenzene (TNCB), and 2,4-dinitrofluorobenzene (Watanabe et al. 2007) at the sensitization phase. In another study, ultraviolet irradiation was shown to activate the NALP3 inflammasome and promote IL-1$\beta$ maturation in keratinocytes (Feldmeyer et al. 2007).

### 1.8.11 NALP3 inflammasome: mechanisms of action

NALP3 is proposed to be involved in the orchestration of both pathogen and danger immune recognition, and it has been suggested that disruption of the ionic environment by NALP3 activators may initiate a cellular alarm signal involved in the activation of the inflammasome (Pétrilli et al. 2007b). The manner in which NALP3 senses these disparate activators is still unknown. Two simplistic scenarios have been proposed (Pétrilli et al. 2007a). The first one suggests a direct ligand-receptor interaction, such as has been proposed for the

TLRs (Pétrilli et al. 2007a). The second model proposes that each ligand could induce a common cellular signal that consequently would become a NALP3 ligand and activator (Pétrilli et al. 2007a).

In line with the second model, Pétrilli *et al.* have recently provided evidence that the activity of NALP3 is blocked by inhibiting potassium efflux, and proposed a model in which the various stresses generated by the different NALP3 agonists would lead to a change in the intracellular ionic milieu, inducing a conformational change in the inflammasome infra-structure and subsequently leading to its activation (Pétrilli et al. 2007b; Carneiro et al. 2008). In their model, the common trigger of NALP3 inflammasome activation would be a decrease in physiological intracellular potassium concentration (Pétrilli et al. 2007b), which is in line with the observation that some of the most potent activators of NALP3 inflammasome (e.g. ATP and nigericin) are known to act through a potassium efflux-dependent mechanism (Perregaux et al. 1994; Walev et al. 1995; Colomar et al. 2003). Furthermore, the inflammasome inducer ATP acts via the $P2X_7$ receptor, which upon stimulation facilitates an immediate potassium efflux (Pétrilli et al. 2007b). Several groups also provided evidence that potassium efflux is required for all other known activators of NALP3 (Fernandes-Alnemri et al. 2007; Franchi et al. 2007; Pétrilli et al. 2007b), and at cytoplasmic concentrations below approximately 100 mM, spontaneous NALP3 activation has been reported to occur (Martinon et al. 2002). Importantly, the activation of the inflammasome by low concentrations of potassium is specific to NALP inflammasomes (Carneiro et al. 2008).

It is now hypothesized that, through the activation of $P2X_7$, ATP released from necrotic cells could induce the opening of potassium channels, resulting in potassium efflux in neighboring cells, although this last point as not been formally demonstrated (Mariathasan et al. 2006). Therefore, it is possible that NALP3 would be able to integrate both host-derived and microbe-derived stimuli, supporting the hypothesis that, rather than being a direct sensor of pathogen, NALP3 could signal "downstream" of an initial sensing event mediated by a separate receptor (Fritz et al. 2006).

## 1.9 Inhibitor of the NALP3 inflammasome: pyrin protein/*MEFV* gene

Noteworthy, the NALP3 inflammasome activity can be dampened down by a variety of cytoplasmic proteins, including pyrin (also known as marenostrin) (Chae et al. 2006), a protein encoded by the *MEFV* gene (The International FMF consortium 1997; The French FMF Consortium 1997). Indeed, pyrin has been shown to regulate caspase-1 activation and consequently IL-1$\beta$ release (Aksentijevich et al. 1999).

Although pyrin is not part of the NLR/CATERPILLER family, it shares the pyrin (i.e. PYD) domain found in NALPs (Ting et al. 2006). Moreover, both NALP3 (previously known as cryopyrin) and pyrin share a similar expression pattern, are involved in similar pathways (i.e. caspase-1 inflammasome pathway, IL-1 regulation, NF-κB pathway, and apoptosis), and mutations in these genes

result in similar clinical presentations with relatively short episodes of systemic inflammation, fever, and the potential development of amyloidosis (Stojanov et al. 2005; Ting et al. 2006).

### 1.9.1 *MEFV* mutations & familial Mediterranean fever (FMF)

The *MEFV* locus mapping to the short arm of chromosome 16p13.3 is mutated in the Familial Mediterranean Fever (FMF) (The International FMF consortium 1997; The French FMF Consortium 1997). FMF is an autosomal, recessively inherited, systemic, auto-inflammatory disease, and is the first HPFS ever discovered. It is commonly found in the Mediterranean basin and in the Middle East (Kastner et al. 2005).

Two positional cloning consortia identified the gene in 1997 using polymorphic microsatellite markers and SNPs spanning the chromosome 16p region. They identified the presence of FMF-carrier haplotypes among several affected populations (The International FMF consortium 1997; The French FMF Consortium 1997). Notably, subgroups of these haplotypes had sets of shared markers that all mapped to the same genomic region, indicating that the carrier chromosomes were derived from a limited number of common ancestors (The International FMF consortium 1997; The French FMF Consortium 1997). Both groups localized the FMF-causing locus in a small overlapping genomic region. The original mutational screens of the 781 amino acid full-length pyrin protein identified four different disease-associated conservative missense mutations in the C-terminal B30.2 domain, which were not present in chromosomes from

ethnically matched non-carrier individuals (The International FMF consortium 1997; The French FMF Consortium 1997).

## 1.9.2 The clinical aspect of the familial Mediterranean Fever

### 1.9.2.1 Prevalence of FMF

In the vast majority of affected families, the disease occurs in members of one generation, consistent with recessive transmission (Bhat et al. 2007). It is the most common HPFS with a prevalence of 0.1% in people of Mediterranean descent (Tucan et al. 2005), with more than 100,000 individuals living with FMF worldwide (Drenthet al. 2001). The disease is most prevalent among non-Ashkenazi Jews, Arabs, Turks and Armenians, with carrier frequencies of 1:5 to 1:16, 1:5, 1:5, and 1:7, respectively (Bhat et al. 2007). However, it is observed worldwide due to the extensive population movements of the 20$^{th}$ century (Kastner et al. 2005).

### 1.9.2.2 Symptoms of FMF

Attacks of FMF consist of unprovoked 1-3 day episodes of high fever with severe abdominal or chest pain, monoarticular arthritis, or an erysipeloid rash (Ting et al. 2006). These attacks are characterized histologically by a massive influx of polymorphonuclear leukocytes into the affected region(s), neutrophilia, and a rapid acute-phase response (Ting et al. 2006). A subset of patients may develop systemic amyloidosis, which is caused by the deposition of a misfolded

fragment of serum amyloid A (an acute-phase protein), and this can sometimes lead to kidney failure and death (Falk et al. 1997).

#### 1.9.2.3 FMF treatment

The use of colchicine in FMF over the past three decades has dramatically changed the course of the disease. Colchicine, an alkaloid, originally extracted from plants of the genus Colchicum, has been used in the treatment of gout since the first century (Bhat et al. 2007). In 1972, Goldfinger first described its effectiveness in preventing FMF attacks, by reporting his experience with colchicine in five patients (Goldfinger 1972). Only 10% of patients are colchicine resistant and continue to suffer attacks at the same intensity as before (Lidar M et al. 2007). Before the advent of colchicine therapy, amyloidosis was reported to occur in about 75% of FMF patients over the age of 40 (Sohar et al. 1967).

One interesting finding was that pyrin interacts with tubulin and co-localized with microtubules (Mansfield et al. 2001), and this would certainly suggest a rationale for the current treatment of the disease that uses colchicine, a microtubule-destabilizing agent.

### 1.9.3 Pyrin expression

Pyrin is expressed at high levels in neutrophils, eosinophils, monocytes and dendritic cells, but not in lymphocytes (The International FMF consortium 1997; The French FMF Consortium 1997; Centola et al. 2000; Diaz et al. 2004). These are all key inflammatory effector cells, further supporting the role of pyrin

in the regulation of the innate immune response. Pyrin is also expressed in synovial, peritoneal and skin-derived fibroblasts, but not in chondrocytes or endothelial cells (Diaz et al. 2004). Moreover, the subcellular localization of pyrin has been reported to vary; in transfection studies, the full-length pyrin localized to the cytoplasm, while a rarer isoform lacking exon 2 entered the nucleus (Chen et al. 2000; Papin et al. 2000; Tidow et al. 2000; Mansfield et al. 2001; Diaz et al. 2004). Additionally, native pyrin has been reported to be cytoplasmic in monocytes and predominantly nuclear in granulocytes, dendritic cells, and synovial fibroblasts (Diaz et al. 2004). Two members of the 14.3.3 protein family, which are known to be involved in the regulation of intracellular signaling, have been shown to bind pyrin through phosphorylated serine residues that are encoded by the exon 2 (Jéru et al. 2005). Mutants at these residues were reported to be unable to bind the 14.3.3 proteins and to enter the nucleus (Jéru et al. 2005). These latter results might partly account for the observed differences in pyrin localization in the different cell types. Noteworthy, the pyrin-14.3.3 protein interaction has not been demonstrated for the endogenous protein yet.

### 1.9.4 Pyrin domains & location of FMF causative mutations

Although the N-terminal half of pyrin seemed to be unique when *MEFV* was initially cloned over 10 years ago, the 92 amino acid N-terminal region was subsequently recognized as the prototype for the motif that bears its name (Bertin et al. 2000; Harton et al. 2002). Pyrin is now considered to be a member of the death fold superfamily (Bertin et al. 2000; Martinon et al. 2001), with its N-

terminal pyrin domain (PYD) homologous to the death domain (DD), death effector domain (DED), and caspase recruitment domain (CARD) subfamilies (Fesik 2000). Importantly, all these domains are known to participate in highly specific homotypic protein-protein interactions and are found in several proteins involved in the regulation of inflammation and apoptosis (Wang et al. 2002; Manji et al. 2002). Pyrin is part of a larger family termed the TRIpartite Motif (TRIM) proteins that typically have a RING, B-Box and coiled-coil (CC) domains, and also frequently the B30.2 domain (Masters et al. 2006a).

The pyrin protein contains at least four different conserved domains, comprising an N-terminal pyrin (i.e. PYD) domain, a B-box zinc-finger, a coiled-coil (CC), and a 160-170 amino acid B30.2 domain (also known as the PrySpry domain) located at the C-terminal segment (Bhat et al. 2007).

Although over 80 mutations in the *MEFV* gene have been described, the majority of FMF cases are caused by four mutations all clustered in exon 10 (i.e. M694V, V726A, M680I and M694I), the prevalence of which varies according to the population studied (Lidar et al. 2007; Bhat et al. 2007). Moreover, over half of the documented missense mutations causing FMF are located in this exon, encoding for the motif B30.2 domain, and this bias is even higher when taking into account the high frequency of individual exon 10 mutations among patients (Milhavet et al. 2008).

### 1.9.5 The biological function of pyrin

Pyrin is thought to be implicated in the regulation of systemic inflammatory response through leukocyte apoptosis, NF-κB transcription factor activation, and IL-1$\beta$ production (McDermott et al. 2007). The biological function of pyrin is the subject of intensive research, and although the precise role of pyrin in the regulation of inflammation in healthy individuals and in the disease state remains unclear, consensus regarding its function is gradually being achieved (McDermott et al. 2007).

Recent studies have provided insights into pyrin's complex interactions with numerous regulatory proteins (Ryan et al. 2008b). The N-terminal PYD domain was found to interact with the PYD domain of the adaptor protein ASC, also involved in the NALP3 inflammasome (Richards et al. 2001). When this interaction was initially described, it was the apoptotic effect of ASC that was addressed. Indeed, the wild-type pyrin, when overexpressed in HeLa cells, appeared to increase ASC "speck" formation, and paradoxically, increased the survival of these cells (Richards et al. 2001; Mansfield et al. 2001). Now, because of the known involvement of ASC in the NALP3 inflammasome complex formation, this interaction suggests a more direct route to the inflammatory hallmarks of FMF.

Both inhibitory and enhancing effects have been observed depending on the experimental design. Experiments performed on animal models and cell lines, as well as the presumed structure of pyrin and the presence of certain protein domains, suggest that the wild type isoform of the protein has anti-inflammatory

properties (McDermott 2004; Chae et al. 2003). Indeed, targeted disruption of pyrin in the mouse resulted in increased susceptibility to endotoxin toxicity and exhibited increased macrophage caspase-1 activation and IL-1$\beta$ processing (Chae et al. 2003). Based on their results, Chae and colleagues proposed that wild-type pyrin could be implicated in the reduction of inflammation (Chae et al. 2003). However, since the endogenous mouse *MEFV* exon 10 is not highly conserved with the analogous human exon, these mice experiments could not capture the full spectrum of pyrin functionality and subsequent studies in human cells were needed (Chae et al. 2000).

### 1.9.6 Pyrin & NALP3 inflammasome

The presence of a common PYD domain in pyrin and NALPs, as well as the similarities between the syndromes associated with the mutations in pyrin and NALP3 suggested that these proteins could be involved in related pathways. In agreement with this hypothesis, several studies have reported the interaction of pyrin with different NALP3 inflammasome components. Depending on the experimental conditions, two mechanisms have been proposed. The first one involves the direct interaction of the amino-terminal death PYD domain with the intracellular adaptor ASC, which then inhibits the caspase-1 activation as the PYD domain of ASC is no longer available to bind NALP3 (Chae et al. 2003; Yu et al. 2006). The second proposed mechanism involves the binding of the carboxy-terminal B30.2 domain of pyrin to the catalytic domain of caspase-1 (Chae et al. 2006). The B30.2 domain is thought to mediate protein-protein interactions, but the mechanism underlying caspase-1 inhibition is unclear.

Importantly, a large percentage of the FMF-associated mutations are located in the B30.2 domain (Chae et al. 2006).

There is a controversy concerning the function of pyrin in the regulation of the inflammasome as both an inhibiting and an activating function have been postulated (Centola et al. 2000; Chae et al. 2003; Chae et al. 2006; Yu et al. 2006). However, the fact that pyrin mutations are associated with inflammatory responses lends itself to the concept that these mutations might confer a "loss-of-function" phenotype, that pyrin may down-regulate the inflammasome function and consequently, caspase-1 activation and IL1-$\beta$ mediated innate response. There is a large body of data indicating an important role for pyrin in the regulation of caspase-1 and IL-1$\beta$ activation that may be context dependent (Chae et al. 2003; Chae et al. 2006; Yu et al. 2006; Papin et al. 2007; Seshadri et al. 2007.), mediated both through homotypic interaction of the N-terminal PYD with ASC and through the inhibitory effects of the C-terminal B30.2 domain on caspase-1 and other components of the inflammasome (Chae et al. 2006; Papin S et al. 2007).

Noteworthy, the B30.2 domain at the C-terminal end of the protein pyrin has been reported to interact with NALP3, ASC, caspase-1, and the inflammasome substrate pro-IL-1$\beta$ (**Figure 1.3**) in human cells, thereby inhibiting the activity of the NALP3 inflammasome and production of IL-1$\beta$ (Richards et al. 2001; Chae et al. 2006; Papin et al. 2007). Moreover, whereas a pyrin knock-down resulted in increased caspase-1 activation and IL-1$\beta$ secretion, overexpression of the B30.2 domain alone was shown to block these processes (Papin et al. 2007). These observations led Papin and colleagues to conclude that

pyrin could bind to several inflammasome components, thereby modulating their activity (Papin et al. 2007). Chae *et al.* also reported that mutations in this domain prevented pyrin from interacting with and inhibiting caspase-1 (Chae et al. 2006). Moreover, the structure of the B30.2 domain has recently been determined (Woo et al. 2006; Masters et al. 2006b; Grütter et al. 2006) and confirmed the possible model of interaction of pyrin and caspase-1 computationally. The modeled complex highlights that key residues mutated in FMF patients, such as the ones at positions M694 and M680, were located proximally to the potential binding interface, suggesting that mutations at such residues could prevent binding of pyrin (Chae et al. 2006). Furthermore, human TPH-1 monocytic leukemic cells treated with Pyrin siRNA or shRNA exhibited increased release of IL-1$\beta$ *in vitro* upon stimulation of the NALP3 inflammasome (Chae et al. 2006; Papin et al. 2007), which is similar to the effects reported in the Pyrin-deficient mice (Chae et al. 2003) discussed previously.

Nevertheless, the overall effect of pyrin on the inflammasome response is still a matter of dispute and may even be context dependent. One study reported that a pyrin molecule containing FMF-associated B30.2 mutations showed reduced caspase-1 interaction and decreased efficiency in blocking IL-1$\beta$ release (Chae et al. 2006). Moreover, using *in vitro* transfection assays, another study by Yu and colleagues reported that overexpression of pyrin in 293T cells that stably express ASC and pro-caspase-1 appeared to activate the NALP3 inflammasome complex, although mutant forms of the pyrin protein did not result in higher levels of ASC-dependent caspase-1 activation (Yu JW et al. 2006). This latter group concluded that pyrin acted as a pro-inflammatory molecule through the formation

of a pyrin inflammasome complex. Although these results seem to be in direct contradiction with other reports previously mentioned, all these experiments have been performed using overexpression systems with different cell types; thus these results may not be comparable.

### 1.9.7 Pyrin & NF-κB activation

In previous reports, the impact of full-length pyrin on NF-kB activation has ranged from an inhibitory effect (Dowds et al. 2003; Masumoto et al. 2003), to a potentiating effect (Sarkar et al. 2006), to a context dependent effect (Stehlik et al. 2002), to no effect at all (Yu et al. 2006). Chae *et al.* recently published a study that may clarify this issue as they found that only N330, but not C330 or full-length pyrin, enhances NF-κB activation, raising the possibility that discrepancies in the literature may have resulted from differing experimental conditions that may or may not favor pyrin cleavage (Chae et al. 2008). Indeed, Chae and colleagues reported that pyrin is a caspase-1 substrate that is cleaved at Asp330, a site that is distal to the B30.2 domain that binds caspase-1 (Chae et al. 2006; Papin et al. 2007). Additionally, they show that pyrin isoforms harboring FMF-associated B30.2 mutations were cleaved more efficiently than wild-type pyrin (Chae et al. 2008). Furthermore, they reported that the N-terminal cleaved fragment interacted with the p65 subunit of NF-kB and with IκBα through its 15 amino acid bZIP basic domain and adjacent sequences (Chae et al. 2008). These latter interactions enhanced the entry of p65 into the nucleus and induced calpain-mediated degradation of IκBα, thus potentiating NFκB activation (Chae et al.

2008). They also demonstrated the nuclear localization of N330, but not full-length pyrin or the C-terminal (C330) fragment, in their transfection system and showed that the N330 (N-terminal) fragment included a site that, when phosphorylated, could bind members of the 14.3.3 family and caused cytoplasmic retention (Jéru et al. 2005; Chae et al. 2008). Interestingly, the absolute and relative quantities of cleaved pyrin and IκBα degradation products were substantially increased in leukocytes from FMF patients compared with healthy controls.

Hence, the differential sensitivity of mutant and wild type pyrin to cleavage of caspase-1 and binding to NF-κB players suggested a pathway by which pyrin mutations could lead to heightened potential for inflammation through NF-κB. Overall, by demonstrating that caspase-1 initiated a pyrin-dependent NF-κB activation pathway, Chae *et al.* established another link between caspase-1 and auto-inflammatory disease (Chae et al. 2008).

### 1.9.8 FMF & IBD

IBD and FMF share common clinical and biologic features, such as the periodicity and relapse of clinical features, infiltration by neutrophils at the site of injury, and an abnormal regulation of apoptosis (Lichtenberger et al. 2004; McDermott 2004).

A number of published epidemiological and genetic studies have reported that IBD is more common and severe in patients with FMF, and some studies suggested that mutations in *MEFV* responsible for FMF could have potential

modifying effect in IBD patients (Cattan et al. 2000; Fidder et al. 2002; Fidder et al. 2005; Karban et al. 2005; Giaglis et al. 2006). Amongst these studies, Cattan *et al.* reported that an FMF and IBD association in non-Ashekazi Jews was 8- to 14- fold higher than expected (Cattan et al. 2000). The authors suggested that the genes responsible for one disorder could have a modifying effect on the other inflammatory disease. Fidder *et al.* reported that CD appeared to be more prevalent in FMF patients and presented later in patients without FMF (Fidder et al. 2002). FMF in this latter group of patients showed a higher attack frequency and was more often complicated by amyloidosis (Fidder et al. 2002). Karban *et al.* also reported that the prevalence of E148Q (*MEFV* exon 2) amongst CD and controls was similar, but that E148Q was more common in CD patients with perianal disease (Fidder et al. 2005). Sari *et al.* also reported the concurrent manifestation of UC and FMF in 3 infants less than 6 months of age, in whom infantile UC was severe, intractable, and associated with a *MEFV* mutation (Sari et al. 2008). All three patients had a M694V mutation in exon 10 (Sari et al. 2008). Based on this clinical, epidemiological, and biological evidence, we hypothesized that the *MEFV* gene could also contribute to inflammatory bowel disease susceptibility (Chapters 2-3).

## 1.10 Irritable bowel syndrome (IBS)

Changes in the composition of the intestinal bacteria microbial flora has also been speculated to be associated with another inflammatory disease of the GI tract called irritable bowel syndrome (IBS). Emerging data suggests that fecal microbial flora of individuals with IBS differs from that of healthy controls (Swidsinki et al. 1999). Moreover, increased levels of anaerobic *Bacteroides* organisms and *Escherichia coli* have also been reported in the mucosa of IBS patients (Whitehead et al. 1988).

### 1.10.1 IBS definition

IBS is considered a GI sensory and motility disorder characterized by a triad of abdominal pain, bloating, and changes in bowel habit, with an absence of any apparent mucosal abnormality (Thompson et al. 1999). Although only recently defined, reports of IBS appeared in the literature back in the early 19$^{th}$ century (Powell 1818). The first modern concept of IBS was proposed by Chaudhary and Truelove in the 1960s (Chaudhary et al. 1962) that characterized many of the key symptoms, including the observation that approximately a quarter of patients' diarrhea began with an enteric infection. Today, IBS is diagnosed by the presence of GI symptoms as outlined by the Rome III criteria together with the exclusion of other organic diseases such as IBD, microscopic colitis, chronic enteric infection (e.g. giardiasis), and celiac disease (DuPont 2008). The Rome III criteria includes the presence of recurrent abdominal pain or

discomfort lasting over three months, with onset of the symptoms present at six months prior to clinical presentation and diagnosis, together with at least two of the following: improvement of symptoms with defecation, onset associated with a change in frequency of stool, and/or onset associated with a change in appearance of stool (Drossman 2006). IBS is further subdivided into four groups based on the predominant stool consistency according to the Bristol stool-form scale: constipation-predominant IBS, diarrhea predominant IBS, mixed or alternating diarrhea-constipation IBS, and unclassified IBS (Thompson et al. 1999; Longstreth GF et al. 2006).

### 1.10.2 IBS: a burden of the health care system

IBS is the most frequent reason for referral to gastroenterology outpatient clinics and has a profound impact on patients' quality of life (Bommelaer et al. 2004). IBS represents a significant burden on the health-care system with patients visiting their physicians 8-10 times a year, and an approximate total of 5 million prescriptions a year in the United States (Sandler et al. 2002; Cash et al. 2005). The costs associated with IBS are high not only for the individual but also for the healthcare system, as illustrated by studies reporting that direct and indirect costs of diagnosis and symptom management of IBS were estimated to be $1.66 billion (Sandler et al. 2002; Shih YT et al. 2002; Cash et al. 2005; Maxion-Bergemann et al. 2006; Shaheen et al. 2006). Additionally, IBS has been shown to reduce the quality of life of patients, affecting the rates of individuals missing work and school (Leong et al. 2003; Cain et al. 2006).

### 1.10.3 IBS pathophysiology: unknown risk factors

Despite its high prevalence, the exact pathogenesis and clinical manifestations leading to IBS-like symptoms remains complex and only partially understood. Hereditary and environmental factors are likely thought to play a role (Levy et al. 2001; Saito et al. 2008a) and many studies have reported abnormal GI motility (i.e. muscle contractions), visceral hypersensitivity, psychological dysfunction, and emotional stress in IBS patients (Ohman et al. 2007). Yet, despite intensive investigations, no specific physiological and psychological abnormality has been found to be absolutely indicative of this disorder (Ohman et al. 2007), and the precise molecular mechanism(s) underlying IBS remain essentially unknown. IBS is a difficult diagnosis to make and can be even more challenging to treat. This has led to imprecise drug development and unsatisfactory patient management, and, together with the high cost associated with treating IBS patients, there is a renewed interest and motivation into studying and determining the underlying pathogenic mechanisms that contribute to IBS with the hopes of identifying better diagnostic tests and therapeutic targets. Within this context, there has been a growing interest in the genetic association study approach, which can be considered a good strategy to assess pathways potentially involved in the development of the disease.

### 1.10.4 IBS genetics

Recent investigations support a genetic susceptibility basis for the development of IBS, even though evidence suggests it plays a modest role and that several genes are likely to be involved, as is the case with complex traits.

#### 1.10.4.1 IBS: fundamental genetic epidemiology

IBS is highly prevalent and has been reported to affect between 10-15% of the American and Western European population (Russo MW et al. 1999; Creed et al. 2001). Population studies have shown that IBS seems to be related to gender and age, but not to ethnicity (Chang et al. 2006). Indeed, studies reported that IBS is most commonly diagnosed between the ages of 15-44 years, that the proportion of women diagnosed and treated with IBS is 2-3 times that of men, but that the proportions of white subjects and non-white subjects affected with IBS are almost equal (Russo et al. 1999; Chang et al. 2006).

#### 1.10.4.2 IBS: familial aggregation

IBS patients often report that family members have similar symptoms, and recent studies support the observation that IBS clusters in families (Kalantar et al. 2003; Locke et al. 2000; Kanazawa et al. 2004). Additionally, children of parents with IBS are twice as likely to seek medical help for abdominal complaints compared to children of parents without IBS (Levy et al. 2000). Yet, a major limitation of most of these studies is that the IBS status of relatives was collected from the patients but not verified with the relatives. In a study where relatives

were directly queried regarding their bowel symptoms, it was reported that between 54-68% of patients with IBS had an affected first-degree family member, compared with 19-36% of controls without IBS (Saito et al. 2008b).

#### 1.10.4.3 IBS: twin study

Five twin studies reported a genetic contribution to IBS ranging between 0-20% (Morris-Yates et al. 1998; Levy et al. 2001; Mohammed et al. 2005; Bengtson et al. 2006; Lembo et al. 2007). The large variability between these studies may be the result of methodological differences with respect to how IBS status was defined and how the information was collected. It is worth mentioning that the British study only included patients who had confirmed IBS according to Rome II criteria (Mohammed et al. 2005), whereas the other studies, including the American (Levy et al. 2001) and Australian studies (Morris-Yates et al. 1998), were limited by the fact that they did not use any consensus criteria to diagnose IBS. Four of these studies performed additional modeling, looking at the correlations between members of a twin pair to resolve whether the observed phenotypic variance could be attributed to environmental or genetic components. Except for the British Twin Study (Mohammed et al. 2005), the three other studies found that the best-fit models included components for additive genetic components and individual environmental exposures. The genetic heritability was estimated to be 22%, 48%, and 57% in the American (Levy et al. 2001), Norwegian (Bengtson MB et al. 2006), and Australian (Morris-Yates et al. 1998) twin studies, respectively, whereas the individual environmental contributors

ranged from 43-78%. Importantly, shared environment did not appear to be an important predictor of phenotypic variance in any of these three studies. Moreover, these studies reported that IBS did not result from a dominant genetic effect. Although the results of these twin studies are revealing, they cannot conclusively prove that IBS did not aggregate between twins due to a specific shared household risk factor like diet, lifestyle, or infection. Overall, these studies suggest that both hereditary and environmental risk factors may contribute to the development of IBS.

#### 1.10.4.4 IBS: genetic association studies

In the past few years, there has been a growing interest in identifying the role of genetic factors in the development of IBS. Several groups have evaluated specific candidate polymorphisms selected based on the biological plausibility that the encoded protein product could play a role in IBS pathogenesis.

The gene that has received the most attention from multiple investigators is the serotonin transporter gene (*SLC6A4*; solute carrier family 6, member 4), encoding for a protein located on the presynaptic terminal that is responsible for re-uptake of serotonin from the synaptic cleft. A study showed that mice lacking this gene exhibited an altered colonic motility similar to IBS-like symptoms (Chen et al. 2001). Recent drug therapies for IBS targeted the serotoninergic pathway, motivating several groups to focus on studying the serotonin transporter gene (Van Kerkhovenet al. 2007; Camilleri 2007). This gene maps to chromosome 17q11.1-q12 (Ramamoorthy et al. 1993) and its most widely studied

variant, 5-HTTLPR, is a 43-base pair insertion-deletion polymorphism located in the promoter region. The short transcript isoform has been associated with lower transcriptional efficiency than the long transcript isoform (Heils et al. 1996; Greenberg et al. 1999). To date, the results of studies examining the relationship between these alleles and specific symptoms of IBS have been inconsistent, and a recent meta-analysis has concluded that this polymorphism is neither associated with IBS, nor with its subtypes (Van Kerkhoven et al. 2007).

Beside this candidate gene, other studies have also evaluated variants in the 5-$HT_{2A}$ receptor gene (*HTR2A*, 5-hydroxytryptamine receptor 2A) (Pata et al. 2004), the norepinephrine transporter gene (*SLC6A2*) (Kim et al. 2004), the $alpha_{2A}$-adrenergic receptor gene (*ADRA2A*) (Kim et al. 2004), the $alpha_{2C}$-adrenergic receptor gene (*ADRA2C*) (Kim et al. 2004), the cholecystokinins (CCK) (Cremonini et al. 2005), the beta-3 subunit G protein (*GNB3*) (Saito et al. 2007; Andresen et al. 2006), as well as some inflammatory genes including IL-10 (*IL10*) (Gonsalkorale et al. 2003; van Der Veek et al. 2005), transforming growth factor-beta1 (*TGFB1*) (Gonsalkorale et al. 2003), and tumor necrosis factor-alpha (*TNF*) (van Der Veek et al. 2005). Positive associations have been reported for the -1082G/A (rs1800896) polymorphism located in the promoter of *IL10* (Gonsalkorale et al. 2003) and for the promoter variant -308G/A (rs1800629) located in the *TNF* region (van Der Veek et al. 2005).

### 1.10.4.5 IBS: challenges of genetic association studies

None of these findings has been replicated in subsequent studies and thus, no definitive conclusions have been drawn regarding these candidates. In the absence of an identified genetic risk factor for IBS, the major obstacle to gene discovery is the lack of additional clinical and epidemiological information regarding IBS. It would greatly benefit study design to clearly identify the environmental risk factors involved in the development of the disease that will need to be adjusted for in future IBS genetic studies. Indeed, well-defined phenotypes are crucial as combining individuals that are clinically and etiologically heterogeneous will decrease the chances of detecting modest genetic risk factors involved in IBS. One option would be to first evaluate candidate genes in more unique group of patients, such as post-infections IBS (PI-IBS) patients. Moreover, control subjects should be drawn from the same population as the cases (e.g. population-based or hospital-based) and should be collected in a similar time frame. Matching on age, gender, and ethnicity is also important because temporal trends, gender-specific behaviors or exposures, or population stratification may influence the disorder (Tabor et al. 2002). Keeping in mind the aforementioned key factors for designing an optimal genetic association study, we have opted to perform a candidate gene association study using the well-defined PI-IBS population cohort derived from the Walkerton community, which will be further described bellow (Chapter 5).

# 1.11 Post-infectious irritable bowel syndrome

## 1.11.1 PI-IBS definition

Most people with IBS describe a subtle onset of bowel symptoms over a period of time, but there is a subgroup of individuals with previously normal bowel habits who describe an acute onset of persistent symptoms after an episode of gastroenteritis characterized by at least two of the following: fever, vomiting, diarrhea, or positive stool culture (Parry et al. 2005). After meeting criteria for IBS (with Rome III criteria being the most recently defined) (Longstreth et al. 2006) in the absence of alarm symptoms and after excluding other causes of persistent bowel dysfunction in the form of diarrhea, urgency, loose stools, and abdominal pain, such patients are said to have post-infectious irritable bowel syndrome (PI-IBS) (Spiller 2003a). Typically, the acute infectious symptoms of vomiting and fever resolve after several days with resolution of the infection; however, abdominal discomfort, bloating, and diarrhea (the predominant bowel alteration in PI-IBS patients) persist.

This condition, like other non-specific forms of IBS, also represents a significant burden on the health-care system, with the majority of patients consulting their family physician within 6 years of symptoms onset, and around 80% requiring further investigation, such as colonoscopy or barium enema (Near et al. 2002). Although the idea of IBS developing after gastrointestinal infection is not new, being first described in 1962 (Chaudhary et al. 1962), the scientific study of the mechanisms involved is relatively recent. PI-IBS appears to be a

nonspecific response to infection caused by a variety of enteric pathogens, and has been documented after illness due to *Campylobacter*, *Salmonella*, diarrheagenic strains of *Escherichia coli,* and *Shigella* species (Spiller et al. 2000; Gwee et al. 2003; Wang et al. 2004). These numerous pathogenic organisms known to be responsible for acute bacterial gastroenteritis contribute in altering the composition of the gut microbial flora (Spiller 2007a), which may result in increasing the intestinal immunological reactivity and lead to the development of IBS.

### 1.11.2 PI-IBS incidence

Acute infectious gastroenteritis is a common event in western countries, occurring on average 1.4 times per year per person (Barbara et al. 2007a). Although the vast majority of subjects recover completely following eviction of the infectious agent, a small significant proportion of subjects go on to develop long-lasting PI-IBS digestive symptoms (Spiller 2003a; Parry et al. 2003; Barbara et al. 2004a). The risk of developing IBS following acute enteric infection has been established based on reports of outbreaks in western nations and in travelers to developing countries, with a 4%–32% reported incidence of PI-IBS (Neal et al. 1997; Neal et al. 2002; Parry et al. 2003; Ilnyckyj et al. 2003; Okhuysen et al. 2004; Borgaonkar et al. 2006; Marshall et al. 2006; Halvorson et al. 2006; Spiller et al. 2006). This wide range of reported incidence rates could be explained by the considerable variability in study design (including in the duration of follow-up), the definition of gastroenteritis, and the criteria used for diagnosing IBS.

Importantly, most studies lacked control groups, which would have made it possible to compare the incidence of IBS in an otherwise similar population without preceding infection. Among the studies that included control groups, the incidence of IBS after acute bacterial enteritis has been reported to vary between 10-15% (DuPont 2008).

A meta-analysis of eight studies by Halvorson and colleagues reported a mean prevalence of IBS of 9.8% among patients with a history of infectious gastroenteritis, compared with a mean prevalence of IBS of 1.2% among control subjects (Halvorson et al. 2006). Thabane and colleagues, who also recently conducted a systematic review and meta-analysis, concluded that the pooled incidence of developing IBS following an intestinal infection was 10% (Thabane et al. 2007), and that the individuals most at risk for developing IBS were younger, more anxious or depressed, and had prolonged fever during infection.

### 1.11.3 Walkerton outbreak of acute bacterial gastroenteritis

One of the largest reported incidences of PI-IBS (36%) occurred after the Walkerton outbreak of acute bacterial gastroenteritis (Marshall JK et al. 2006). Walkerton is a small rural town located in a prime agricultural area 180 kilometers Northwest of Toronto, Canada. In May 2000, public municipal water was contaminated with *Escherichia coli* 0157:H7 and *Campylobacter* species. Heavy rainfall contributed to surface transport of livestock fecal contaminants into inadequately chlorinated drinking water (Marshall JK et al. 2006). This is

also the largest prospective study conducted to date; the outbreak produced an estimated 2300 cases of acute gastrointestinal illness, over 750 emergency room visits, 65 admissions, 27 cases of hemolytic uremic syndrome, and 7 deaths (Marshall JK et al. 2006).

The Walkerton Health Study (WHS) was initiated to study the epidemiology and long-term health outcomes of the waterborne outbreak of acute gastroenteritis, and facilitate local residents' access to specialty clinical care. There were 904 cases of self-reported gastroenteritis and several documented positive stool cultures. After a 2 year follow-up, the incidence of IBS was 36.2% in residents who had gastroenteritis and met Rome I criteria compared to 10.2% of this who did no, representing and OR of 4.8 (95% CI 3.4-6.8; $p<0.001$) (Marshall JK et al. 2006). The predominant symptom was diarrhea and independent risk factors included younger age and female gender (Marshall JK et al. 2006).

This unfortunate tragedy provided a unique opportunity to study the genetic determinants of PI-IBS (Chapter 5) in the large, well-defined Walkerton study cohort with simultaneous and well-characterized exposure (Marshal JK. Gastroenterology 2006).

### 1.11.4 PI-IBS risk factors

A cohort study that used a large U.K. primary care database identified acute bacterial gastroenteritis as the strongest risk factor identified to date for the

development of IBS (adjusted relative risk 11.9) (Rodríguez et al. 1999). Other known risk factors involved in the development of IBS following an enteric infection include female gender, psychological factors, duration and severity of the acute illness, younger age, and microbial virulence factors (Thabane et al. 2007; Spiller 2007b; DuPont AW 2008). However, genetic factors had never been evaluated as potential risk factors contributing to PI-IBS susceptibility (Chapter 5).

### 1.11.5 PI-IBS genetic association study: defining the three axes of candidate genes

Patients with PI-IBS appear to have chronic mucosal immunologic dysregulation with altered intestinal permeability and motility that can lead to persistent intestinal symptoms (DuPont 2008). It has been hypothesized that PI-IBS patients have an inability to down-regulate the initial inflammatory stimulus to the gut, resulting in increased levels of serotonin-producing enterochromaffin cells, T lymphocytes, macrophages, and pro-inflammatory cytokines (Barbara et al. 2007b). Moreover, animal studies support the hypotheses that intestinal inflammation (Adam et al. 2006; La et al. 2004; Bercik et al. 2004; Akiho et al. 2005), alterations in gut flora (Verdu et al. 2006), and intestinal permeability (Ferrier et al. 2003), resulting in an activated intestinal immune system, may explain part of the gut dysfunction observed in PI-IBS patients.

In considering these aforementioned risk factors for PI-IBS, we selected candidate genes that had been characterized and potentially associated with IBS,

the serotoninergic pathway, the innate immune response, and the intestinal epithelial barrier function. To select the final candidates, we proceeded on the basis of a comprehensive overview of the intestinal inflammatory system biology, and with the hypothesis that PI-IBS might result from persistent local low-grade inflammation. The three main axes of research will be briefly reviewed.

#### 1.11.5.1 Axis I: Serotonin & serotoninergic pathway

Five percent of the serotonin (5-hydroxytryptamine) in humans is concentrated in the central nervous system and 95% is in the GI tract within enterochromaffin cells, neurons, mast cells, and smooth muscle cells (Gershon 1999). When serotonin is released, it affects vagal afferent nerve fibers and enteric nerve fibers, resulting in intestinal secretions and peristalses. These physiological responses to serotonin release may result in the symptoms such as nausea, vomiting, pain, bloating, and altered bowel habits (Gershon 1999).

More recently, drugs acting at the serotonin receptor subtypes have been developed, tested, and marketed for the management of specific subgroups of patients with IBS. The first serotonin agent that has been reported to be effective in reducing stool frequency and urgency in women with diarrhea-predominant IBS was alosetron, a 5-$HT_3$ receptor antagonist (Camilleri et al. 2000; Whorwell 2001; Camilleri 2007).

Serotonin specifically released from enterochromaffin cells affects GI motililty, enterocyte secretion, and visceral sensation (Coates et al. 2004). A 25% increase in the number of rectal enterochromaffin cells has been documented in

patients with PI-IBS, compared to patients with non-PI-IBS and control subjects (Dunlop et al. 2003a; Dunlop et al. 2003b). In addition, an increase in post-prandial plasma serotonin levels has been observed in patients with PI-IBS compared with healthy control subjects (Dunlop et al 2005).

In addition to focussing on serotonin, we also selected candidates involved in the serotoningergic pathway known to influence the production and recognition of serotonin. We also included any polymorphisms previously evaluated in published IBS genetic association studies.

### 1.11.5.2 Axis II: Inflammatory changes

Although the definition of IBS is based on the lack of structural and biochemical abnormalities that can be identified using routine diagnostic tests (Thompson et al. 1999), there is accumulating evidence supporting the activation of the intestinal immune system as a risk factor for the development of IBS (Thornley et al. 2001; Dunlop et al. 2003a; Dunlop et al. 2003b).

Animal studies have highlighted the roles of severity and transient mucosal inflammation in terms of specific abnormalities of GI functions (Liebregts et al. 2005; Adam et al. 2006). These observations led research groups to speculate that IBS may, in at least a subpopulation, be an inflammatory disorder, keeping in mind that the inflammatory component may very well be different in PI-IBS relative to IBS patients without an onset related to infection.

Immunohistological studies have revealed mucosal immune system activation in a subset of IBS patients, mostly in those with diarrhea-predominant

type (Chadwick et al. 2002). Similar observations have been made in patients with presumed PI-IBS, suggesting a possible common link. Moreover, increased concentrations of intra-epithelial lymphocytes (Spiller et al. 2000; Chadwick et al. 2002), mast cells (Weston et al. 1993; O'Sullivan et al. 2000; Chadwick et al. 2002; Guilarte et al. 2006), and 5-hydroxytryptophan-secreting enterochromaffin cells (Spiller et al. 2000) have been observed in the mucosa of IBS patients, with particularly increased levels in PI-IBS patients (Spiller et al. 2000). Mast and enterochromaffin cells are both specialized cells that are part of the innate immune system, helping to clear pathogens by stimulating propulsive gut motility (Spiller 2007b).

The inflammatory component of IBS has been more clearly demonstrated in patients with an infectious onset of the disease. Studies have shown that patients with PI-IBS have an activated intestinal immune response as defined by increased numbers of T cells in the colonic mucosa (Spiller et al. 2000; Dunlop et al. 2003a; Dunlop et al. 2003b) and an augmentation of mucosal enteroendocrine cell levels (Spiller et al. 2000), which may suggest that there is a continuing inflammatory stimulus (Spiller et al. 2000). Furthermore, the increase in intra-epithelial lymphocyte and lamina propria lymphocyte counts have been reported to persist for at least a year after infection in PI-IBS patients, and have been associated with increased intestinal permeability (Spiller et al. 2000). Studies showed that the mRNA expression levels of the pro-inflammatory cytokine IL-1$\beta$ were significantly increased in rectal biopsies from PI-IBS patients (Spiller et al. 2000; Wang et al. 2004). However, no studies have clearly demonstrated whether

the inflammatory components of PI-IBS are driving or are the result of the clinical symptoms of the patients.

Transient mucosal inflammation as well as the role of the severity of inflammation are now believed to be an important trigger for the manifestation of IBS-like symptoms (Liebregts et al. 2005), as shown by both animal (Liebregts et al. 2005; Adam et al. 2006) and human (Thornley et al. 2001; Spiller 2003b) studies. It is possible that the control of the inflammatory response could be compromised in some individuals due to genetic predisposition, and that patients who are exposed to a severe GI infection in the presence of specific genetically determined risk factors might have a greatly increased risk of developing PI-IBS. Thus, for the second group of candidates, we opted to focus on genes involved in the inflammatory immune response.

#### 1.11.5.3 Axis III: Mucosal permeability

The gut microbial flora is physically separated from the host by a thin intestinal epithelial barrier. "Tight junctions" between the cells of the intestinal barrier are involved in regulating the transport of molecules while prohibiting the migration of microorganisms (Turner 2006). This selective permeability can be altered directly by bacteria, or indirectly via cytokines produced by the host immune response to a microbial challenge (Nusrat et al. 2000). Increased intestinal permeability may play an important role because it exposes the mucosa to luminal antigens from dietary and bacterial origins, promoting and maintaining mucosal immune activation (Barbara et al. 2005). Moreover, the enhanced neuromuscular exposure to these antigens may also lead to altered visceral

sensitivity and enteric dysmotility through chronic inflammatory mechanisms (DuPont 2008).

Three independent groups have reported that PI-IBS patients have an increased small intestinal permeability (Spiller et al. 2000; Marshal et al. 2004; Dunlop et al. 2006). One of these groups is the Walkerton Health Study, which reported that 35% of the new PI-IBS patients had increased intestinal permeability compared to just 13% in the non-IBS controls (Marshal et al. 2004). Moreover, one study showed that non-PI-IBS patients with predominant diarrhea symptoms had a significant increase in small intestinal permeability when compared to PI-IBS patients and healthy controls (Dunlop et al. 2006). Overall, increased permeability resulting from a breakdown of the normal intestinal barrier allows access of bacterial products to the lamina propria and this may be one of the mechanisms contributing to the perpetuation of chronic inflammation. For the third group of PI-IBS candidates, we targeted genes involved in maintaining the integrity of the intestinal epithelial barrier.

# 2 INVESTIGATION OF THE *NLRP3* GENE IN CROHN'S DISEASE

Reproduced with permission from *Nature Genetics*.

Full copy of the published letter can be found in appendices.

**Reference:**

Villani AC, Lemire M, Fortin G, Louis E, Silverberg MS, Collette C, Baba N, Libioulle C, Belaiche J, Bitton A, Gaudet D, Cohen A, Langelier D, Fortin PR, Wither JE, Sarfati M, Rutgeerts P, Rioux JD, Vermeire S, Hudson TJ, Franchimont D. Common variants in the NLRP3 region contribute to Crohn disease susceptibility. *Nat Genet* 2009; 41: 71-76.

**Affiliations of co-authors:**

Alexandra-Chloé Villani[1,2], Mathieu Lemire[3], Geneviève Fortin[1], Edouard Louis[4], Mark S Silverberg[5], Catherine Collette[1], Nobuyasu Baba[6], Cécile Libioulle[4], Jacques Belaiche[4], Alain Bitton[7], Daniel Gaudet[8], Albert Cohen[9], Diane Langelier[10], Paul R Fortin[11], Joan E Wither[11], Marika Sarfati[6], Paul Rutgeerts[12], John D Rioux[13], Severine Vermeire[12], Thomas J Hudson[3] & Denis Franchimont[1,14]

[1] Division of Gastroenterology, McGill University Health Centre, Montreal General Hospital, 1650 Cedar Avenue, Montréal, Québec H3G 1A4, Canada; [2] McGill University and Génome Québec Innovation Centre, 740 Doctor Penfield Avenue, Montréal, Québec H3A 1A4, Canada; [3] Ontario Institute for Cancer Research, MaRS Centre, South Tower, 101 College Street, Suite 800, Toronto, Ontario M5G 0A3, Canada; [4] Division of Gastroenterology, CHU, Université de Liège, 13 avenue de l'Hôpital Bât. B35, 4000 Liège, Belgium; [5] Mount Sinai Hospital IBD Centre, University of Toronto, 441-600 University Avenue, Toronto, Ontario M5G 1X5, Canada; [6] Immunoregulation Laboratory, CHUM Research Center, Université de Montréal, 1560 rue Sherbrooke Est, Montréal, Québec H2L 4M1, Canada; [7] Division of Gastroenterology, McGill University Health Centre, Royal Victoria Hospital, 687 Pine Avenue West, Montréal, Québec H3A 1A1, Canada; [8] Université de Montréal, Centre de médecine génique communautaire de l'Université de Montréal, Hôpital universitaire de Chicoutimi, 555 boulevard de l'Université, Chicoutimi, Québec G7H 2B1, Canada; [9] Jewish General Hospital, 3755 chemin de la Côte-Sainte-Catherine, Montréal, Québec H3T 1E2, Canada; [10] Centre Hospitalier Universitaire de Sherbrooke–Hôtel-Dieu, 580 rue Bowen Sud, Sherbrooke, Québec J1G 2E8, Canada; [11] Arthritis Centre of Excellence, Toronto Western Hospital Research Institute, 399 Bathurst Street, University of Toronto, Toronto, Ontario M5T 2S8, Canada; [12] Department of Gastroenterology, University Hospital Gasthuisberg, 49 Herestraat, 3000 Leuven, Belgium; [13] Université de Montréal and Montreal Heart Institute, Research Center, 5000 rue Bélanger, Montréal, Québec H1T 1C8, Canada; [14] Department of Gastroenterology, Erasme Hospital, ULB, 808 Lennik road, 1070 Brussels, Belgium.

## 2.1 Rationale

For the following study, the selection of the *NLRP3* gene, enconding the NALP3 protein, as a candidate gene for CD susceptibility was based on the knowledge of the important role played by the microbial flora and bacterial sensors in the onset of CD, the known *NOD2* association with CD susceptibility, the role of NOD2 as an intracellular bacterial sensor (PRMs), and the fact that mutations within the NOD2 NACHT domain that are associated with Blau syndrome are structurally associated with the NALP3 mutations responsible for three auto-inflammatory disorders (i.e. MWS, FCAS, and NOMID). Additionally, NALP3 is part of the same CATERPILLER gene family as NOD2 (Ting et al. 2006), is known to also play a role in intracellular bacterial sensing like NOD2 (Ting et al. 2006), and was reported to detect MDP, the same ligand recognized by NOD2 (Martinon et al. 2004). Moreover, mutations in both *NOD2* and *NLRP3* are known to lead to rare auto-inflammatory syndromes. Together with the known biological paralleles between NALP3 and NOD2, we hypothesized that while severe mutations in these genes will cause the respective monogenic syndrome, milder genetic changes such as SNPs may give rise to more subtly altered transcript, and hence contribute to the susceptibility of chronic inflammatory diseases like CD. We thus evaluated the contribution of *NRLP3* to CD susceptibility.

# 2.2 Common variant in *NLRP3* region contribute to Crohn's disease susceptibility

## 2.2.1 Abstract

We used a candidate gene approach to identify a set of SNPs, located in a predicted regulatory region on chromosome 1q44 downstream of *NLRP3* (previously known as *CIAS1* and *NALP3*), that are associated with Crohn's disease. The associations were consistently replicated in four sample sets from individuals of European descent. In the combined analysis of all samples (710 father-mother-child trios, 239 cases and 107 controls), these SNPs were strongly associated with risk of Crohn's disease ($P_{combined} = 3.49 \times 10^{9}$, odds ratio = 1.78, confidence interval = 1.47–2.16 for rs10733113), reaching a level consistent with the stringent significance thresholds imposed by whole-genome association studies. In addition, we observed significant associations between SNPs in the associated regions and *NLRP3* expression and IL-1$\beta$ production. Mutations in *NLRP3* are known to be responsible for three rare auto-inflammatory disorders. These results suggest that the *NLRP3* region is also implicated in the susceptibility of more common inflammatory diseases such as Crohn's disease.

## 2.2.2 Introduction

Crohn's disease and ulcerative colitis are multigenic and heterogeneous inflammatory bowel diseases of the gastrointestinal tract that seem to result from a dysregulated mucosal immune response to bacterial antigens in the gut lumen of a genetically susceptible host (Podolsky et al. 2002). *NLRP3* is a member of the CATERPILLER (Ting et al. 2006) family of genes encoding for proteins that comprise a nucleotide-binding domain and a leucine-rich repeat domain. NALP3 (also known as cryopyrin), the protein encoded by *NLRP3*, controls the inflammasome, a crucial molecular platform that regulates activation of caspase-1 and processing of IL-1$\beta$, two key mediators of inflammation (Mariathasan et al. 2007; Pétrilli et al. 2007a; Martinon et al. 2007). The potential involvement of *NLRP3* in the pathogenesis of more common inflammatory disorders motivated us to conduct an in-depth genetic analysis of the *NLRP3* region.

## 2.2.3 Methods

**Subjects.** Five sample sets from four different centers, totaling 710 Crohn's disease trios, 239 Crohn's disease cases and 107 controls, were assembled for this project (**Table 2.1**). Informed consent was obtained from all participants, and protocols were approved by the local Institutional Review Board of all participating Institutions. A clinical subtype of Crohn's disease was assigned using standard clinical criteria (Lennard-Jones1989; Silverberg et al. 2005), except for a few 'indeterminate colitis' cases that were excluded from the study. Belgian subjects from Leuven University Hospital Gasthuisberg (Leuven) were used for the exploratory experiments ($n$ = 296 trios). The replication cohorts

consisted of Belgian subjects from Université de Liège and of two Canadian cohorts (Québec and Toronto). Samples from Liège were subdivided into a family-based cohort (155 Liège trios) and a case-control cohort (Liège case-control, 239 Crohn's disease and 107 controls). The Québec cohort ($n$ = 130 trios) comprised subjects from multiple sites in the province of Québec and included 22 probands of Ashkenazi and 5 of Sephardic Jewish ancestry. The Toronto samples ($n$ = 129 trios) were collected from multiple sites in Toronto and included 26 probands of Ashkenazi Jewish ancestry. All study participants were of European descent except for 15 probands from Toronto that were excluded from the analyses.

All patients were recruited through specialized hospitals, academic centres, and practitioners. Inflammatory bowel disease (IBD) specialists involved in this study confirmed the diagnosis of Crohn's disease and patients were excluded from the study in the case of doubtful diagnosis. In all the participating centres, the diagnosis of IBD was made by IBD specialists after fulfilling standard clinical, radiological, endoscopic, and pathological criteria (Lennard-Jones 1989) that required (1) one or more of the following symptoms: diarrhea, rectal bleeding, abdominal pain, weight loss, fever or complicated perianal disease; (2) occurrence of symptoms on two or more occasions in the past or ongoing symptoms of at least 4-6 weeks' duration; (3) evidence of inflammation, strictures or fistula from radiological, endoscopic, and histological evaluation (with some specific Crohn's disease characteristics); (4) exclusion of all other diagnoses

besides Crohn's disease. Only subjects of European ancestry were used in the final analysis.

**Genotyping.** We first investigated a 67.8-kb region spanning 1q44 (243890897–243958709; NCBI build 35, hg17), including *NLRP3* (32.9 kb). Samples were genotyped using the SNPstream ultra-high-throughput genotyping system (Orchid Biosciences) (Bell et al. 2002) and Sequenom homogenous MassExtend (hME) assays (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) (van den Boom et al. 2007). Primers and probes are available in **Supplementary Table 2.1**. Analysis was restricted to SNPs passing quality filters, excluding SNPs with success rate < 95%, minor allele frequency < 5% or deviation from Hardy-Weinberg equilibrium ($P < 0.01$). Measures of linkage disequilibrium between SNPs and departures from Hardy-Weinberg equilibrium were computed using Haploview v4.0 (Barrett et al. 2005). Families showing mendelian errors in 5% or more of the markers were excluded from the analysis (25 Crohn's disease families).

**Sequencing.** Primers were designed to have a $T_m$ of 60 °C using the Primer3 program (**Supplementary Table 2.2**). PCR reactions were done using Hot Start Taq polymerase (Qiagen) in an 8-μl final volume comprising 9 ng of DNA (one cycle of 96 °C for 10 min and 40 cycles of 95 °C for 30 s, 58 °C for 1 min, 72 °C for 35 s and one cycle of 72 °C for 7 min). Sequencing was done on an ABI 3730 DNA sequencer (Applied Biosystems) according to standard protocols. Sequence

traces were assembled and analyzed using the PolyPhred software package and were compared to annotated sequences from NCBI build 35, hg17.

**SNP selection for the exploratory and follow-up genotyping phases**. We first investigated a 67.8kb region spanning 1q44 (243890897-243958709) (NCBI Build 35, hg17), including *NLRP3* (32.9kb). 43 tagging SNPs were selected using data from HapMap Public Release #22 ($r^2 \geq 0.8$) (minor allele frequency (MAF) $\geq 0.05$) and 14 additional SNPs were selected from dbSNP Build126 in regions with lower coverage. Following the 9kb resequencing experiment, 13 SNPs located within *OR2B11* coding, conserved or predicted regulatory regions were prioritized (**Supplementary Table 2.3**). For the 58 SNPs not previously genotyped, we used a pairwise tagging approach (de Bakker et al. 2005) ($r^2 \geq 0.8$) to select tagging SNPs that best captured the variation in the region (**Supplementary Table 2.3**). A total of 24 SNPs were genotyped in the Leuven exploratory Crohn's disease trios, and 15 of them were analysed after excluding 4 rare SNPs (MAF <0.05), 2 with a low success rate (<95%), and 3 not in Hardy-Weinberg equilibrium ($p < 0.01$).

**Statistical analysis.** Tests of association were done using the likelihood methods implemented in UNPHASED v3.0.10 (Dudbridge et al. 2008), which can analyze samples of nuclear families, unrelated subjects or a combination of both. For nuclear families (cases and their parents), the likelihood is equivalent to the conditional likelihood models on which the transmission disequilibrium test is

based (Spielman et al. 1993). For samples of unrelated cases and controls, the likelihood is equivalent to a logistic model, which allows estimations of risk effects in terms of odds ratios along with their confidence intervals. It also allows conditioning on the observed association at one marker, to test whether or not the observed significance at others can solely be explained by linkage disequilibrium. A permutation procedure is implemented to allow the calculation of significance levels that are corrected for the number of tests. Attributable and prevented fractions were calculated under the assumption that alleles had additive effects on the penetrance scale, that odds ratios from **Table 2.1** are good approximations for the relative risks and that allele frequencies in the controls are reasonable estimates of population frequencies.

**Tissue collection**. *Animal*: Male BALB/c mice (6-8 weeks) from Charles River Laboratory (St-Constant, Quebec, Canada) were maintained under conventional housing conditions. Acute colitis was induced by rectal instillation of 2.5mg of trinitrobenzene sulfonate (TNBS) (Sigma Aldrich Canada Ltd., Ontario, Canada) and mice were sacrificed on day 2 or day 4 (Neurath et al. 1995). Chronic colitis was induced by a secondary instillation of TNBS on day 7 and mice were sacrificed 4 days later. All mice were handled according to institutionally recommended animal care guidelines and all experiments were approved by the Animal Studies Ethics Committee of McGill University. *Human*: Colonic or ileal biopsy specimens were obtained from Crohn's disease patients and from patients undergoing colorectal cancer screening (i.e. healthy controls). All participants

gave informed consent and the study was approved by the Institutional Review Board.

**Monocyte isolation and culture**. Peripheral blood cells (PBCs) were isolated by density gradient centrifugation from healthy volunteers' whole blood. Collected cells were further purified by cold aggregation and rosetting with sheep's red blood cells. Monocyte-enriched PBCs were plated at a concentration of $3X10^6$ cells/ml in serum-free RPMI 1640 (Invitrogen, Ontario, Canada), and non-adherent cells were removed after 1 hour of incubation at 37°C. Adherent cells were cultured for 3 hours in the presence or absence of lipopolysaccharide (LPS) (E.Coli, Sigma-Aldrich, 1$\mu$g/ml) in RPMI 1640 +10% fetal bovin serum (FBS) (Invitrogen). Upon harvest, culture supernatants were collected for cytokine assessment, and monocytes were lysed in RLT buffer (Qiagen RNeasy Mini Kit).

**RNA extraction and quantitative real-time PCR.** Biopsies preserved in RNAlater (Qiagen) were homogenized, and total RNA was extracted using TRIzol (Invitrogen). Total RNA was extracted from monocytes using an RNeasy kit (Qiagen). Total RNA was extracted from PBCs using a PAXgene blood RNA kit (Qiagen) with RNase inhibitor, using off-column DNase I digestion and ethanol precipitation to improve RNA yield and quality. First-strand cDNA was synthesized from 1 μg of RNA template with a cDNA archive kit (Applied Biosystems), using MultiScribe reverse transcriptase and random primers. Quantitative real-time PCR assays (mouse *Nlrp3*, Mm00840904_m1; human

*NLRP3*, Hs00918082_ml; 18S RNA, 4319413E; Applied Biosystems) were conducted using an ABI PRISM 7900 sequence detection system based on the 5' nuclease assay (Holland et al. 1991) and quantified using Applied Biosystems' comparative threshold cycle ($C_t$) method. The Wilcoxon signed-rank test was used to evaluate tissue (mouse and human) expression differences. Associations between *NLRP3* expression or IL-1$\beta$ level and the genotypes of the Crohn's disease–associated SNPs were assessed using linear regression. Analyses were done using GraphPad Software.

**DNA extraction.** Monocyte DNA was isolated from whole blood using a FlexiGene DNA kit (Qiagen). PBC DNA was isolated using a Gentra Autopure automated system (Qiagen) according to the manufacturer's protocol.

**Enzyme-linked immunosorbent assay.** IL-1$\beta$ was quantified in monocyte culture supernatants using a human IL-1β DuoSet ELISA kit (R&D Systems) according to the manufacturer's protocol.

## 2.2.4 Results

We first assessed the association between 47 SNPs in the *NLRP3* region and Crohn's disease risk in 296 trios from Leuven University Hospital Gasthuisberg (see Methods and **Supplementary Table 2.1**). The major alleles of three tagging SNPs were significantly associated with increased risk of Crohn's

disease in the Leuven sample set ($P = 0.0107$ for rs4353135; $P = 7.63 \times 10^3$ for rs4266924; $P = 1.68 \times 10^3$ for rs10733113; **Figure 2.1** and **Table 2.1**). These SNPs span a 5.3-kb region and are located 4.7 kb downstream of *NLRP3* and 1.85 kb upstream of the olfactory receptor gene *OR2B11* (**Figure 2.1**). No associations were observed with SNPs within the *NLRP3* gene, and none of the associated SNPs were in linkage disequilibrium with tagging SNPs located in *NLRP3* or *OR2B11*. These preliminary associations pointed to the 5.3-kb region near *NLRP3*, bounded by rs4353135 and rs10733113, as a candidate region contributing to Crohn's disease susceptibility.

In the second phase of the study, we examined the association of the above three SNPs with risk of Crohn's disease in additional samples. Overall, we screened one case-control cohort (Liège) and three familial sample sets (Liège, Québec and Toronto; see Methods). We replicated our initial significant Crohn's disease associations with the major alleles of tagging SNPs rs4266924 and rs10733113 in all four sample sets ($P < 0.05$; **Figure 2.1** and **Table 2.1**). For the Québec and Toronto cohorts, the observed associations remained whether the analysis was done with or without samples from individuals of Jewish ancestry (data not shown). Association with rs4353135 was replicated only in the Toronto sample set ($P < 0.05$; **Figure 2.1** and **Table 2.1**). Combined analysis of all Crohn's disease samples revealed strong associations for rs4353135 ($P_{combined} = 8.36 \times 10^3$, odds ratio = 1.21, confidence interval = 1.05–1.39, T allele frequency = 71% in cases and 65% in controls), rs4266924 ($P_{combined} = 6.01 \times 10^7$, odds ratio = 1.69, confidence interval = 1.37–2.07, A allele frequency = 91% in cases and

84% in controls) and rs10733113 ($P_{combined}$ = 3.49 × $10^9$, odds ratio = 1.78, confidence interval = 1.47–2.16, G allele frequency = 90% in cases and 80% in controls).

This regional association was the first step in localizing the most likely causal variant(s). Because the observed association signal was not in linkage disequilibrium with any genotyped variant within *NLRP3*, we resequenced a 9-kb region extending from the *NLRP3* 3' UTR to the 5.3-kb region described above, inclusively, which also comprised *OR2B11* (**Supplementary Table 2.2**). Overall, we selected 16 Crohn's disease samples and 8 controls based on genotypes at markers rs4266924 and rs10733113 to fully define the linkage disequilibrium within the region and identify all polymorphisms in linkage disequilibrium with the associated SNPs. The resequencing effort identified 79 SNPs, 8 of which were previously genotyped. Among these SNPs, 14 were novel compared to dbSNP release 129, and 60 had a minor allele frequency ≥0.05 (**Supplementary Table 2.3**).

We next conducted comprehensive genotyping to identify polymorphisms with stronger associations (Methods). We genotyped a total of 24 SNPs in the Leuven exploratory Crohn's disease trios (**Supplementary Table 2.3**) and analyzed 15 of them after they passed quality control tests (see Methods). None of the SNPs within *OR2B11* (**Supplementary Table 2.3**) were associated with Crohn's disease. The major alleles of three SNPs, spanning a 1.8-kb region bounded by rs4353135 and rs10733113, were associated with Crohn's disease (**Figure 2.1** and **Table 2.1**). According to the computational method ESPERR

(Taylor et al. 2006), two of these SNPs (rs6672995 and rs55646866) are located in a predicted regulatory region. The third variant, ss107635144, was selected by the tagging algorithm (de Bakker et al. 2005). These SNPs were in high linkage disequilibrium ($r^2 > 0.70$) with rs4266924 and rs10733113. None of these three SNPs were tags for the SNPs that were not genotyped in the region.

We subsequently screened these three variants in the other four sample sets, and their associations were consistently replicated (**Figure 2.1** and **Table 2.1**). Combined analysis of all Crohn's disease samples revealed strong associations for rs55646866 ($P_{combined} = 7.2 \times 10^{7}$, odds ratio = 1.69, confidence interval = 1.38–2.08, C allele frequency = 91% in cases and 84% in controls), rs6672995 ($P_{combined} = 2.91 \times 10^{6}$, odds ratio = 1.53, confidence interval = 1.28–1.82, G allele frequency = 87% in cases and 79% in controls) and ss107635144 ($P_{combined} = 8.50 \times 10^{6}$, odds ratio = 1.53, confidence interval = 1.27–1.84, C allele frequency = 88% in cases and 80% in controls).

Conditioning on one of these associated SNPs to evaluate the residual significance of the others did not provide evidence of additive effects of the associated SNPs on the risk. Rather, the association they all showed with Crohn's disease can be explained solely by linkage disequilibrium. To account for multiple testing issues, we note that if all 62 SNPs had been genotyped in the Leuven exploratory sample in the same study phase, then rs55646866 would have reached a significance level ($P = 5.72 \times 10^{4}$) that, after correction using a permutation procedure, would have still been significant ($P_{corrected} = 0.019$, estimated from

5,000 replicates; see Methods). This result was further strengthened by the consistent replication of rs55646866 in all studied samples.

We next evaluated associations between genotype and gene expression to determine whether the above SNP associations reflected *cis*-acting regulatory effects on *NLRP3*. We first assessed the influence of the six SNPs (**Table 2.1**) on *NLRP3* mRNA expression from freshly isolated peripheral blood cells (PBCs) and from monocytes isolated from the peripheral blood, as *NLRP3* is primarily expressed in granulocytes and monocytes (Kummer et al. 2007) (Methods). A significant association was observed between *NLRP3* expression and rs4353135 genotypes in PBCs ($P$ = 0.00246; **Figure 2.2a**) and monocytes ($P$ = 0.0124; **Figure 2.2b**), with homozygosity for the risk allele being associated with the lowest level of *NLRP3* expression. Results for the five other Crohn's disease–associated SNPs are shown in **Supplementary Figure 2.1** and **Table 2.4**.

Because *NLRP3* is involved in IL-1$\beta$ processing, we therefore also evaluated whether these six SNPs influenced IL-1$\beta$ production. We cultured monocytes in the presence or absence of crude lipopolysaccharide (LPS; Methods), as its derivatives have been shown to stimulate *NLRP3* expression (Martinon et al. 2004). We then assessed IL-1$\beta$ levels in culture supernatants. We observed a borderline-significant association between IL-1$\beta$ levels and rs6672995 genotype under the unstimulated condition ($P$ = 0.0502; **Figure 2.2c**) and a significant association under the LPS-stimulated condition ($P$ = 0.00591; **Figure 2.2d**). In both cases, homozygosity for the risk allele was associated with the

lowest level of IL-1$\beta$. Results for the five other Crohn's disease–associated SNPs are shown in **Supplementary Figure 2.2** and **Table 2.5**.

We also examined *Nlrp3* expression in colon tissues isolated from mice with TNBS-induced colitis, a model that mimics Crohn's disease–like intestinal inflammation, and in biopsies from individuals with Crohn's disease. *Nlrp3* expression was significantly higher in both acute (fold change = 9.38 ± 1.58; $P < 0.0009$) and chronic (fold change = 2.70 ± 0.88; $P < 0.0152$) TNBS-induced colitis models than in colon tissues from control mice (**Figure 2.2e**; see Methods). *NLRP3* expression was also significantly higher in the ulcerated intestinal mucosa from human Crohn's disease samples (fold change = 4.08 ± 1.33; $P < 0.0028$) than in healthy controls (**Figure 2.2f**; see Methods).

## 2.2.5 Figures

**Figure 2.1: Association results for the five Crohn's disease sample sets**

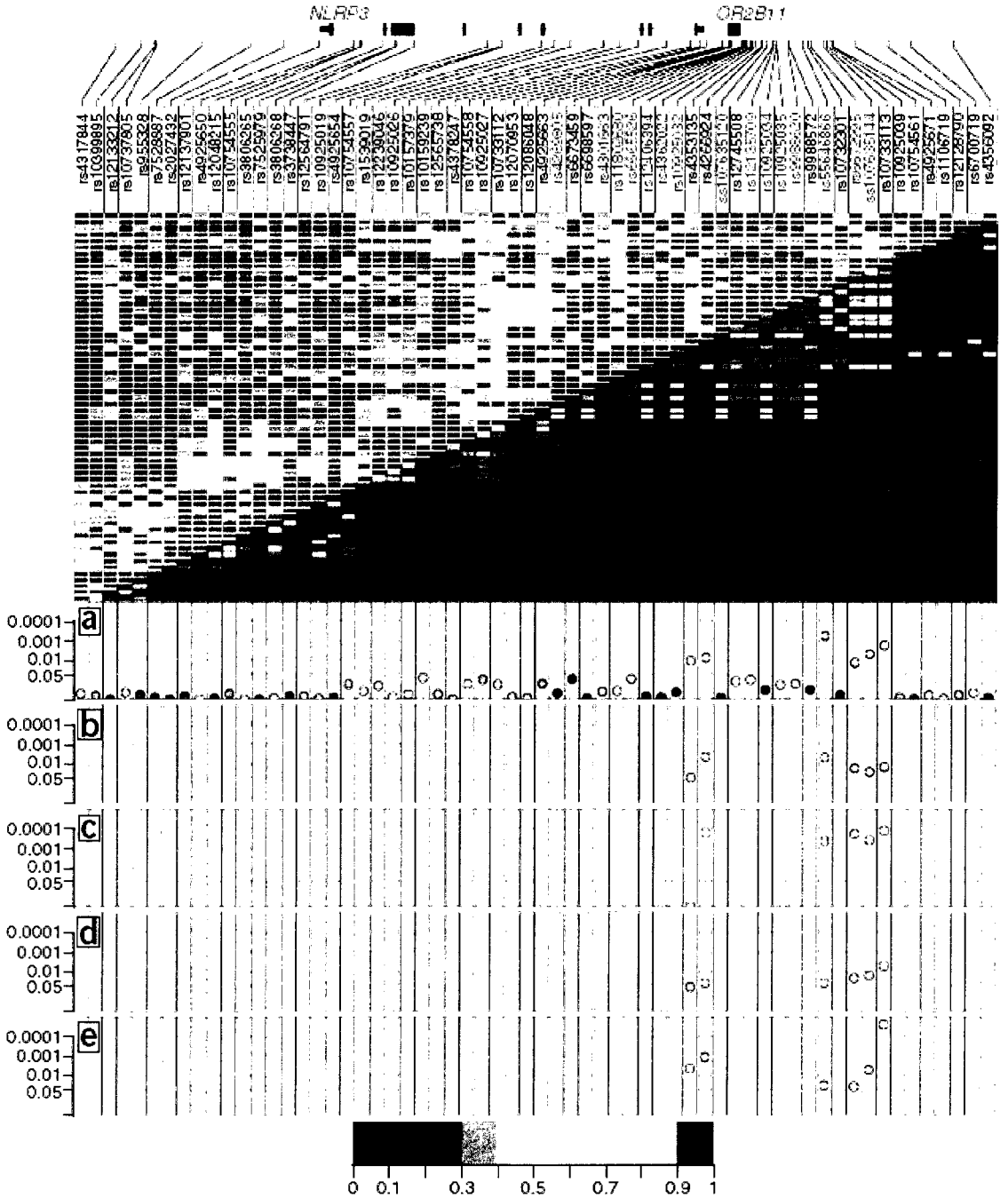


**Figure 2.1:** Top panel shows SNPs, their positions in the genes and the linkage disequilibrium structure between them. SNP names in red were genotyped in the second phase of the study, subsequent to the sequencing experiment. Middle panel shows $D'$ in the upper left and $r^2$ in the lower right. (a–e) Lower panels show results from association analysis of Leuven trios (a), Liège trios (b), Liège case-control cohort (c), Québec trios (d) and Toronto trios (e). $P$ values for individual alleles are reported in a logarithmic scale on the $y$ axis. Color spectrum represents strength of linkage disequilibrium and frequency of the associated alleles.

**Figure 2.2: NLRP3 functional study results**

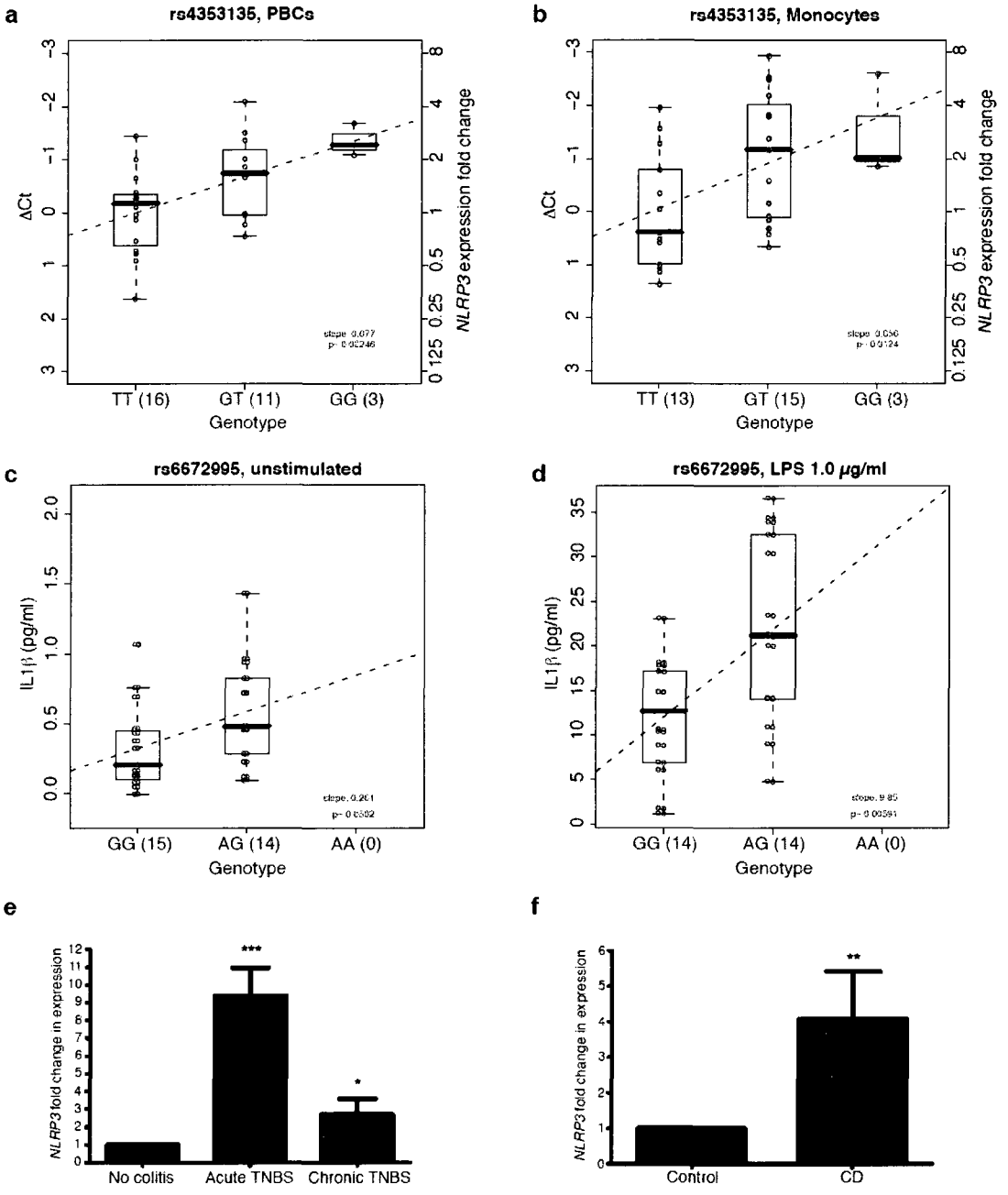


**Figure 2.2:** (a,b) Linear regression analysis of *NLRP3* mRNA level versus rs4353135 genotype in DNA-RNA matched freshly isolated PBCs (a; $n = 30$) and monocytes (b; $n = 31$) obtained from healthy individuals. Genotypes of the six Crohn's disease (CD)-associated SNPs (**Table 2.1**) were obtained by sequencing. Mean threshold cycle ($C_t$) was calculated for each sample from three replicates and then used to calculate relative expression level ($\Delta C_t$), which is the difference between *NLRP3* $C_t$ and endogenous control 18S RNA $C_t$. Fold change in *NLRP3* expression was calculated using comparative $C_t$ method (see Methods), using as a reference the average $\Delta C_t$ of homozygosity for the risk allele. (c,d) Linear regression analysis of IL-1$\beta$ production (pg/ml) versus rs6672995 genotype for unstimulated (c) and LPS-stimulated (d; 1.0 μg/ml) conditions after 3 h of incubation. $\Delta C_t$ (a,b) and IL-1$\beta$ level (c,d) for each individual are shown in red; regression lines are shown as dashed lines (a–d). (e) Quantitative real-time PCR analysis of *Nlrp3* expression in colons of healthy mice ($n = 6$), mice with acute TNBS-induced colitis ($n = 12$) and mice with chronic TNBS-induced colitis ($n = 6$). (f) Quantitative real-time PCR analysis of *NLRP3* expression in colon specimens from healthy individuals ($n = 35$) and individuals with Crohn's disease ($n = 25$). Expression was normalized to 18S RNA expression; each bar represents mean fold change in *NLRP3* expression ± s.e.m. normalized to that of healthy colon specimens (e,f).

## 2.2.6 Table

**Table 2.1: Association between SNPs in the 5.3kb region and risk to Crohn disease**

| SNP | Associated Allele | Frequency Control[a] | Frequency Case[b] | T:U[c] | Control Ratios[d] | Case Ratios[d] | Odds ratio (95% c.i)[e] | *p*-value[e] |
|---|---|---|---|---|---|---|---|---|
| **Leuven (296 trios)** | | | | | | | | |
| rs4353135[f] | T | 0.65 | 0.73 | 118:82 | | | 1.44 (1.08-1.91) | 0.0107 |
| rs4266924[f] | A | 0.86 | 0.91 | 57:31 | | | 1.78 (1.14-2.79) | $7.63 \times 10^{-3}$ |
| rs55646866[g] | C | 0.86 | 0.93 | 57:26 | | | 2.19 (1.37-3.52) | $5.72 \times 10^{-4}$ |
| rs6672995[g] | G | 0.83 | 0.88 | 64:39 | | | 1.64 (1.09-2.48) | 0.0133 |
| ss107635144[g] | C | 0.84 | 0.91 | 64:36 | | | 1.78 (1.16-2.73) | $4.82 \times 10^{-3}$ |
| rs10733113[f] | G | 0.84 | 0.91 | 65:34 | | | 1.91 (1.24-2.95) | $1.68 \times 10^{-3}$ |
| **Liege trios (155 trios)** | | | | | | | | |
| rs4353135[f] | T | 0.62 | 0.73 | 33:20 | | | 1.65 (0.93-2.92) | 0.0727 |
| rs4266924[f] | A | 0.84 | 0.94 | 22:7 | | | 3.14 (1.36-7.26) | $4.31 \times 10^{-3}$ |
| rs55646866[g] | C | 0.85 | 0.95 | 20:6 | | | 3.33 (1.36-8.18) | $4.80 \times 10^{-3}$ |
| rs6672995[g] | G | 0.79 | 0.89 | 31:15 | | | 2.07 (1.11-3.84) | 0.0171 |
| ss107635144[g] | C | 0.80 | 0.90 | 27:13 | | | 2.08 (1.03-4.18) | 0.0253 |
| rs10733113[f] | G | 0.83 | 0.92 | 24:10 | | | 2.40 (1.16-4.96) | 0.0148 |
| **Liege CC (239 CD and 107 controls)** | | | | | | | | |
| rs4353135[f] | T | 0.65 | 0.65 | | 39:44:11 | 92:95:27 | 1.01 (0.71-1.46) | 0.944 |
| rs4266924[f] | A | 0.79 | 0.90 | | 63:35:4 | 184:41:2 | 2.43 (1.53-3.85) | $1.62 \times 10^{-4}$ |
| rs55646866[g] | C | 0.80 | 0.91 | | 66:32:4 | 191:38:3 | 2.33 (1.47-3.69) | $4.32 \times 10^{-4}$ |
| rs6672995[g] | G | 0.74 | 0.86 | | 57:40:7 | 172:56:4 | 2.19 (1.46-3.30) | $1.87 \times 10^{-4}$ |
| ss107635144[g] | C | 0.75 | 0.86 | | 57:38:7 | 165:56:3 | 2.13 (1.41-3.23) | $3.94 \times 10^{-4}$ |
| rs10733113[f] | G | 0.76 | 0.88 | | 58:38:5 | 177:47:3 | 2.36 (1.52-3.65) | $1.20 \times 10^{-4}$ |
| **Quebec (130 trios)[h]** | | | | | | | | |
| rs4353135[f] | T | 0.67 | 0.73 | 68:48 | | | 1.42 (1.00-2.01) | 0.0627 |
| rs4266924[f] | A | 0.85 | 0.91 | 43:26 | | | 1.65 (1.02-2.70) | 0.0400 |
| rs55646866[g] | C | 0.84 | 0.90 | 43:26 | | | 1.65 (1.02-2.70) | 0.0400 |
| rs6672995[g] | G | 0.78 | 0.85 | 53:32 | | | 1.66 (1.09-2.52) | 0.0220 |
| ss107635144[g] | C | 0.80 | 0.88 | 49:28 | | | 1.75 (1.12-2.73) | 0.0160 |
| rs10733113[f] | G | 0.79 | 0.89 | 53:28 | | | 1.89 (1.19-3.00) | $5.10 \times 10^{-3}$ |
| **Toronto (129 trios)[i]** | | | | | | | | |
| rs4353135[f] | T | 0.62 | 0.74 | 63:35 | | | 1.80 (1.17-2.77) | $4.40 \times 10^{-3}$ |
| rs4266924[f] | A | 0.82 | 0.92 | 39:16 | | | 2.50 (1.44-4.35) | $1.12 \times 10^{-3}$ |
| rs55646866[g] | C | 0.83 | 0.90 | 36:20 | | | 1.80 (1.07-3.04) | 0.0313 |
| rs6672995[g] | G | 0.77 | 0.85 | 46:28 | | | 1.64 (1.05-2.56) | 0.0355 |
| ss107635144[g] | C | 0.78 | 0.88 | 46:23 | | | 2.00 (1.22-3.27) | $5.18 \times 10^{-3}$ |
| rs10733113[f] | G | 0.77 | 0.91 | 50:17 | | | 3.00 (1.76-5.10) | $2.47 \times 10^{-5}$ |
| **Combined CD (710 trios, 239 CD and 107 controls)[j]** | | | | | | | | |
| rs4353135[f] | T | 0.65 | 0.71 | 282:185 | 39:44:11 | 92:95:27 | 1.21 (1.05-1.39) | $8.36 \times 10^{-3}$ |
| rs4266924[f] | A | 0.84 | 0.91 | 161:80 | 63:35:4 | 184:41:2 | 1.69 (1.37- 2.07) | $6.01 \times 10^{-7}$ |
| rs55646866[g] | C | 0.84 | 0.91 | 156:78 | 66:32:4 | 191:38:3 | 1.69 (1.38-2.08) | $7.20 \times 10^{-7}$ |
| rs6672995[g] | G | 0.79 | 0.87 | 194:114 | 57:40:7 | 172:56:4 | 1.53 (1.28-1.82) | $2.91 \times 10^{-6}$ |
| ss107635144[g] | C | 0.80 | 0.88 | 186:100 | 57:38:7 | 165:56:3 | 1.53 (1.27-1.84) | $8.50 \times 10^{-6}$ |
| rs10733113[f] | G | 0.80 | 0.90 | 192:89 | 58:38:5 | 177:47:3 | 1.78 (1.47-2.16) | $3.49 \times 10^{-9}$ |

**Table 2.1 footnotes**: [a]: Frequencies in controls are estimated from untransmitted alleles or unrelated controls; [b]: Frequencies in cases are estimated from transmitted alleles or unrelated cases; [c]: Ratio of transmitted versus unstransmitted alleles from heterozygous parents; [d]: The genotype distribution in the cases and controls as x/y/z: homozygous for the high-risk allele (x), heterozygous (y), homozygous for the second allele (z); [e]: Odds ratios, confidence interval and *p*-values are computed from a likelihood test that assumes a multiplicative model for the risk, as implemented in UNPHASED (Dudbridge et al. 2008) (Methods). Allele association was evaluated in the case-control cohort. All tests are two tailed; [f]: SNPs associated with CD in exploratory phase; [g]: SNPs associated with CD in second genotyping phase; [h]: Includes 27 CD trios of Jewish ancestry; [i]: Includes 26 CD trios of Jewish ancestry; [j]: Includes 53 CD trios of Jewish ancestry.

### 2.2.7 Discussion

*NLRP3* (chromosome 1q44) encodes NALP3, which is involved in the inflammasome signaling platform by regulating caspase-1 activity and IL-1$\beta$ processing. The importance of NALP3 in inflammation is highlighted by gain-of-function mutations within its NOD domain that are associated with three hereditary periodic fever syndromes: Muckle-Wells syndrome, familial cold autoinflammatory syndrome and neonatal-onset multisystem inflammatory disease (Agostini et al. 2004; Mariathasan et al 2007). Hyperproduction of IL-1$\beta$ is thought to be a central event leading to symptoms in these three syndromes (Agostini et al. 2004). Consistent with these observations is the successful use of IL-1$\beta$ targeted therapy for treating Muckle-Wells syndrome (Hawkins et al. 2003) and familial cold autoinflammatory syndrome (Hoffman et al. 2004).

In our study, contrary to the gain-of-function mutations described above leading to hyperproduction of IL-1$\beta$, we uncovered a regulatory region downstream of *NLRP3* that contributes to Crohn's disease susceptibility and is associated with hypoproduction of IL-1$\beta$ and decreased *NLRP3* expression. Indeed, the risk allele of rs6672995, located in a predicted regulatory region (Taylor et al. 2006), was associated with a decrease in LPS-induced IL-1$\beta$ production, and the risk allele of rs4353135 was associated with a decrease in baseline *NLRP3* expression in two independent sample sets of healthy donors. It is noteworthy that these two SNPs were in weak linkage disequilibrium ($r^2 <$ 0.105) in our combined sample set. Although the causal variant has not been conclusively shown and may still be unidentified, it is most likely to be in linkage

disequilibrium with the tested variants. Nonetheless, we have shown that SNPs in the associated 5.3-kb region influence *NLRP3* at both the gene expression and functional levels, as indicated by altered *NLRP3* expression and IL-1$\beta$ production. Notably, dysregulated IL-1$\beta$ production has also been linked to Crohn's disease pathogenesis in which the three major *NOD2* (chromosome 16q12) mutations result in a loss-of-function phenotype, with decreased NF-κB activation in response to muramyl dipeptide stimulation and decreased IL-1$\beta$ production in primary human mononuclear and dendritic cells from individuals with Crohn's disease (Li et al. 2004; Van Heel et al. 2005; Kramer et al. 2006; van Beelen et al. 2007). Our results further support the recent Crohn's disease immunopathogenesis paradigm, which suggests that a defective innate immune response impairs clearance of luminal antigens and/or pathogens and leads to the development of chronic intestinal inflammation and Crohn's disease. *NLRP3* may thus have a role in the initiation phase of the disease, as indicated by our *in vitro* and expression experiments, as well as a role in perpetuating chronic inflammation through further activation of caspase-1 and processing of IL-1$\beta$, as indicated by the enhanced *NLRP3* expression in the Crohn's disease and chronic TNBS-induced colitis samples.

The *NLRP3* locus can be added to the list of several newly uncovered Crohn's disease loci at which the common allele has been reported to be the risk allele (Duerr et al. 2006; Rioux et al. 2007; Libioulle et al. 2007; Parkes et al. 2007; Wellcome Trust Case Control Consortium 2007a; Barrett et al. 2008). Although it is difficult to strictly distinguish between one allele being a

susceptibility risk factor and the other being a protective one, estimating attributable fractions and prevented fractions in addition to odds ratios offers insight into how to interpret these associations with very common risk factors in the context of complex diseases. If the SNPs with the strongest associations from **Table 2.1** are interpreted as risk factors, then the attributable fractions (that is, the reduction in prevalence if the risk factor were removed from the population) of the alleles fall in the range of 45–55%. These proportions of 'cases explained' (all other factors being ignored) are large, but not surprisingly so, as most members of the population are carriers of the risk factors. Conversely, if the minor alleles are interpreted as protective factors, then their prevented fractions (that is, the proportion by which the prevalence would increase if the protective factor were removed from the population) fall in the range of 10–14%, a range easier to interpret in the context of complex genetic diseases.

Several recent genome-wide association studies have identified new Crohn's disease susceptibility genes using the Illumina HumanHap300 Genotyping BeadChip (Duerr et al. 2006; Rioux et al. 2007; Libioulle et al. 2007) and the Affymetrix GeneChip Human Mapping 500K Array Set (Wellcome Trust Case Control Consortium 2007a). Although rs4353135 is present on the Human Mapping 500K Array Set, this SNP showed the weakest significance of all six SNPs from **Table 2.1**, was the only one not consistently replicated across all samples and was in weak linkage disequilibrium with the other SNPs (max $r^2$ = 0.28). **Supplementary Table 2.6** shows the linkage disequilibrium between SNPs on the Illumina HumanHap300 and the Affymetrix GeneChip Human Mapping

500K arrays that were not genotyped in the present study and the SNPs from **Table 2.1** found in HapMap. With a maximum $r^2$ of only 0.16, these observations may explain why this region escaped detection in these genome-wide association studies.

A recent meta-analysis of three large genome-wide association studies of Crohn's disease reported that well-established associations with Crohn's disease account for ~20% of the genetic variance observed in Crohn's disease, suggesting that additional genetic contributions have yet to be discovered (Barrett et al. 2008). With the exception of variations within *NOD2* and *IL23R*, established susceptibility alleles have been reported to have relatively modest effects, with odds ratios ranging from 0.7 to 1.7 (Barrett et al. 2008). Despite the modest contribution of *NLRP3* to the risk of Crohn's disease, our results strongly implicate *NLRP3*, a gene with an essential role in regulating the inflammasome, as a risk factor for Crohn's disease. Our results also suggest that a gene such as *NLRP3* that is associated with rare, severe autoinflammatory disorders can also be implicated in the susceptibility of more common inflammatory diseases such as Crohn's disease.

### 2.2.8 Accession numbers

GenBank: *NLRP3*, AF054176; *OR2B11*, NM_001004492. OMIM: *NLRP3*, 606416; *OR2B11*, 605956. SNP data have been submitted to NCBI dbSNP under the numbers ss107635120, ss107635122, ss107635124, ss107635126, ss107635128, ss107635130, ss107635132, ss107635133, ss107635136, ss107635138, ss107635140, ss107635142, ss107635144, ss107635146.

### 2.2.9 Acknowledgement

We thank the study participants and their families for their collaboration; A. Montpetit and the genotyping and sequencing platforms of the McGill University and Genome Québec Innovation Centre for their technical assistance; T. McKenzie, G. Bonventi and C. Landolt-Maticorena for their help with acquisition and preparation of the PBCs; F. Martinon and J. Tschopp for their technical assistance in designing the functional studies; and T. Kwan for helpful discussion. This research was supported by the following grants to D.F.: the Canada Foundation for Innovation, the Canada Research Chair program, the Research Institute of the McGill University Health Centre and the Crohn's & Colitis Foundation of Canada (CCFC). A.-C.V. is supported by a Ph.D. training award from the Fonds de la recherche en santé du Québec. None of the funding agencies had any role in the design and conduct of the study; in the collection, analysis and interpretation of the data; or in the preparation, review or approval of the manuscript.

### 2.2.10 Supplementary tables

**Supplementary Table 2.1: Primers and probes of the 3 genotyping panels used in the study**

| ID | Sequence | Panel ID |
|---|---|---|
| rs4925671-F | AGAGGGAGAGTAACTGCACCA | SNPstream exploratory panel |
| rs4925671-R | ATTTTAATTCATGAGCTACAATTCATG | SNPstream exploratory panel |
| rs4925671-GA-U1 | CAGCACTATTACCATCACGTGGTGCTGCTATTCACAGTGGTTCCC | SNPstream exploratory panel |
| rs10157379-F | AGAGTCCTTGGCAATGGC | SNPstream exploratory panel |
| rs10157379-R | TTAATTGGGAAGATAGAAACCAGT | SNPstream exploratory panel |
| rs10157379-GA-U2 | CTCAGACTACGAATCCACGTGTTTTTTAAGCTCTTTGGTGTCTAC | SNPstream exploratory panel |
| rs12133212-F | TTAAACATTTTGTCCTTTGCTCT | SNPstream exploratory panel |
| rs12133212-GA-R | ATCACAATAAAAGCAGCACAAA | SNPstream exploratory panel |
| rs12133212-GA-U3 | CAGCCATCCATTCACTATCTATGTTCTCCACCACACTGAATGACG | SNPstream exploratory panel |
| rs9988572-F | AACTGGTGGGTGCTCATTC | SNPstream exploratory panel |
| rs9988572-R | TACAAAGGGGTTTGAGGACA | SNPstream exploratory panel |
| rs9988572-GA-U4 | ATCTAACGCACCTACGACCTTACTTCCCTTCTCCTCTCAATGCTA | SNPstream exploratory panel |
| rs10925019-F | CCAACTTCAAGCAAATGGAC | SNPstream exploratory panel |
| rs10925019-R | TAGTGCCATGGAGACTGGTT | SNPstream exploratory panel |
| rs10925019-GA-U5 | CAACAAGTAATCCGCAGACTTTTGGGCATGAAGTTGGAGAAATAA | SNPstream exploratory panel |
| rs10925027-F | CTCCACGATTTTCAATCTCC | SNPstream exploratory panel |
| rs10925027-R | TGGCAAATGATTCCATTGTA | SNPstream exploratory panel |
| rs10925027-GA-U6 | ACAACTCACGCAAGTACCATTTTTCAATCTCCTATCCTATCAAAG | SNPstream exploratory panel |
| rs10733113-F | CAGAACAATGAGTGGCACAG | SNPstream exploratory panel |

| | | |
|---|---|---|
| rs10733113-R | ACTACTTCTTGCGGCCTGTC | SNPstream exploratory panel |
| rs10733113-GA-U7 | TACAAGCACGCACTAGACATGCTTTTTGTTGTTGTTGTTGTTGAT | SNPstream exploratory panel |
| rs6698597-F | CAAACACTCTGACGGTGATG | SNPstream exploratory panel |
| rs6698597-R | TGTGTCTTTGCTCTCCTTCC | SNPstream exploratory panel |
| rs6698597-GA-U8 | AGACTTCTACGCAAGCACTGTTCTCACATACTTAATGTCTCTATC | SNPstream exploratory panel |
| rs10733112-F | TCTCCACAAATTTTACACGACTT | SNPstream exploratory panel |
| rs10733112-R | AGGATCCACAATGTTTTCCA | SNPstream exploratory panel |
| rs10733112-GA-U9 | CCGCCAGTAAGACCTAGACGCCACAAATTTTACACGACTTTTTCA | SNPstream exploratory panel |
| rs2027432_F | ATGGAGACTCAGGTCTGCTG | SNPstream exploratory panel |
| rs2027432_R | CGTGAGGCCTTTAAAACAGA | SNPstream exploratory panel |
| rs2027432_GA_U10 | TCCAGAATAGACAACAGACGCTCTGCAGTTCTGTAGAAAGGTGGT | SNPstream exploratory panel |
| rs10737805-F | TTTCTTTTTATCTGAGTTTGTCGC | SNPstream exploratory panel |
| rs10737805-R | TTCTCTGTGAGATGAAAGAATTTACA | SNPstream exploratory panel |
| rs10737805-GA-U11 | AGCAAGACCACCTAGACCAGCATCATTTCATCTCCTTTGCATCAG | SNPstream exploratory panel |
| rs10802510-F | TGATATTACCCATTTCTTTGCAT | SNPstream exploratory panel |
| rs10802510-R | TGCTAATTGAATCATTCCACA | SNPstream exploratory panel |
| rs10802510-GA-U12 | CCAGATCCTCACCATGTAAGACTACATATTTATGAATTACAATGC | SNPstream exploratory panel |
| rs10754561-F | ATTTAATCACATGAATTGTAGCTCAT | SNPstream exploratory panel |
| rs10754561-R | TTCCTCAGTTGTTTGGCTTAA | SNPstream exploratory panel |
| rs10754561-GA-U13 | CAACAATACGAGCCAGCAAGCTACAAGAATTCTATTTGGCCATAA | SNPstream exploratory panel |
| rs4925650_F | GGAGAATTGCTTGAACATGG | SNPstream exploratory panel |
| rs4925650_R | TTTGCCTCCACCTTCTCATA | SNPstream exploratory panel |
| rs4925650-GA-U14 | AACATACAGACGCACTCCTCGGTTTTTYCTCTTTTCCCCTCTCAC | SNPstream exploratory panel |
| rs947181-F | GGTGGGGTTAGGTTTTAGGT | SNPstream exploratory panel |

| | | |
|---|---|---|
| rs947181-R | CAGCCCATAAAAGACCTCTG | SNPstream exploratory panel |
| rs947181-GA-U15 | ACAATCAACATACGAACAGCTTACTCAGGTGGCTGAAGTGGGTGC | SNPstream exploratory panel |
| rs3806265-F | GGAGCATTTCTGCACTCCTA | SNPstream exploratory panel |
| rs3806265-R | TGTTTGCTGATTGATTTCCA | SNPstream exploratory panel |
| rs3806265-GA-U16 | ATACCTACCACGCTACAGCCTTTCTGCACTCCTAGTTTCAAAACA | SNPstream exploratory panel |
| rs12089794-F | TCCAGCTTGGGTTTCATC | SNPstream exploratory panel |
| rs12089794-R | CAAGATCATGCCACAGCA | SNPstream exploratory panel |
| rs12089794-GA-U17 | GCAGACAACGAACAACTACCAAAAAATTCATCTAGTTCCACCTCA | SNPstream exploratory panel |
| rs10925039-F | GGATGCAGAGGATGGAGTAA | SNPstream exploratory panel |
| rs10925039-R | ACAGGCCGCAAGAAGTAGT | SNPstream exploratory panel |
| rs10925039-GA-U18 | AACATCCACGCAACTCATACGTGGGGTAAAATGTACGTTATTCGG | SNPstream exploratory panel |
| rs4925654-F | TAAGGATAGGGACAGAAAGGAAG | SNPstream exploratory panel |
| rs4925654-R | AACAAAAACTATACTTTATCCCTGTCA | SNPstream exploratory panel |
| rs4925654-GA-U19 | CCACTCAACTCCACGAATACTAACAGGCTCCAGGACCTATGAACT | SNPstream exploratory panel |
| rs955328-F | AGGGACATTTGAGAGCAGTC | SNPstream exploratory panel |
| rs955328-R | CGTACAGTGTGTGTCCCTTG | SNPstream exploratory panel |
| rs955328-GA-U20 | ACAACTACCGACGACAAGACAACCCTGAGGACATTTGCTGAGTGC | SNPstream exploratory panel |
| rs12745508-F | TTAATTCTGGCTGAGTTCCCT | SNPstream exploratory panel |
| rs12745508-R | ATTGTGGTGGGATCACAGC | SNPstream exploratory panel |
| rs12745508-GA-U21 | GATCCATCAACAGACATCACTCCTCATTCACTCTAGTTCAGCACC | SNPstream exploratory panel |
| rs4266924-F | GGGAGAAAAGGGAGAAGTTG | SNPstream exploratory panel |
| rs4266924-R | TGCTGCCGTGTAACTAAGGT | SNPstream exploratory panel |
| rs4266924-GA-U22 | GCAACATAAGACCGCTCAACGCCTTAATATCTCAGCAAGGTCCAT | SNPstream exploratory panel |
| rs12564791-F | AACATTCATTCAGTGCGTGC | SNPstream exploratory panel |

| | | |
|---|---|---|
| rs12564791-R | TAGACCATCAGCTCCCCG | SNPstream exploratory panel |
| rs12564791-GA-U23 | AGTAGCCTAACAGCACTCGAATGTCTGTCCAGGCGTGGCAAAGAT | SNPstream exploratory panel |
| rs3806268-F | ATGATGAGCATTCTGAGCCT | SNPstream exploratory panel |
| rs3806268-R | ATATAGAACAGATAGTCAAACCTGTCTTG | SNPstream exploratory panel |
| rs3806268-GA-U24 | CAACAAGACATAACAACGCACCAGGAAGATGATGTTGGACTGGGC | SNPstream exploratory panel |
| rs10754557-F | AAATCAGAAGTGTACATAGAGCTTGT | SNPstream exploratory panel |
| rs10754557-R | CAACCTTCCATAGAGATGGC | SNPstream exploratory panel |
| rs10754557-GA-U25 | CCATAACAACTTACCAGCCAGCTTCCTTGTCCATGGTGGAGCGTG | SNPstream exploratory panel |
| rs7525979-F | ACATCATCTTCCTGGCCA | SNPstream exploratory panel |
| rs7525979-R | ATTAAGATGGAGTTGCTGTTTGA | SNPstream exploratory panel |
| rs7525979-GA-U26 | AGACCGACAAGCAATCTACATCCCTGCCGCCCCCTGGAACACCAC | SNPstream exploratory panel |
| rs4317844-F | AGAACCTGGGGGTACAAAAG | SNPstream exploratory panel |
| rs4317844-R | CAGTTCCAGACTCTCGCTGT | SNPstream exploratory panel |
| rs4317844-GA-U27 | GCAAGCCATCAGCTAATACAGGGGTACAAAAGGCATCGTGAACCT | SNPstream exploratory panel |
| rs12565738-F | GAGCAACTTGGAGGAAGTGA | SNPstream exploratory panel |
| rs12565738-R | AATTCTGTGTCTCCGTGCTT | SNPstream exploratory panel |
| rs12565738-GA-U28 | CACTACATACGACCGCAGAAAGAATCAGGTTTCGGGTAGAAGAGA | SNPstream exploratory panel |
| rs10159239-F | GAGGAAAGCTGTGTGAGGAG | SNPstream exploratory panel |
| rs10159239-R | TGTCACAGGAGACACAAATGAT | SNPstream exploratory panel |
| rs10159239-GA-U29 | AATAAGCTCACCACCGTCAATCAGTTTGCCTTGGCTCTTTCTGTC | SNPstream exploratory panel |
| rs10732301-F | ATTTTACACTTGCTGCTTAATACGT | SNPstream exploratory panel |
| rs10732301-R | GCAGGATGGGCAGTATTAAC | SNPstream exploratory panel |
| rs10732301-GA-U30 | CAGTCAACAATCCAGATCAACATCGACCCATCCATCCTCCACCTC | SNPstream exploratory panel |
| rs12070953-F | TTTATTTCTCTCTTCTATTCCATAATCAC | SNPstream exploratory panel |

| | | |
|---|---|---|
| rs12070953-R | TACATCTCATGTCCTCATCATCC | SNPstream exploratory panel |
| rs12070953-GA-U31 | CAGAACATCCTCAGAAGCAATTCAYCTACACCCTGAGAAATAAAG | SNPstream exploratory panel |
| rs12086048-F | TGCTCAGGATCCAGTCCTC | SNPstream exploratory panel |
| rs12086048-R | ATGTAAATCGCAGGTAGGTAGAAG | SNPstream exploratory panel |
| rs12086048-GA-U32 | CAAGCAACGACCTACTACAAAGGGACGACACAAGGCCTTTGGGAC | SNPstream exploratory panel |
| rs12048215-F | AGGGTATTGCAGGATGATGA | SNPstream exploratory panel |
| rs12048215-R | CTGCACACTGCTGTCTGTTT | SNPstream exploratory panel |
| rs12048215-GA-U33 | CGCAGAAGCAACTCACTTCTAAGGCAGGGCAGCAGTGTGGGTGTA | SNPstream exploratory panel |
| rs10399895-F | AAGTAGAAGGGTGGGGAGGT | SNPstream exploratory panel |
| rs10399895-R | GGCACATCCTGAATCAGTTC | SNPstream exploratory panel |
| rs10399895-GA-U34 | CACTAGTCATAACGCAGCCTAAGGGTGGGGAGGTGGAGGTTGAAG | SNPstream exploratory panel |
| rs4925663-F | CCTTCTTCGTGTTGGTGC | SNPstream exploratory panel |
| rs4925663-R | TCCCTTGGAGGACTGGAT | SNPstream exploratory panel |
| rs4925663-GA-U35 | CAGAATAGCCACGCCTAGATCTGGCTCTCATCCTTCTCTCCTATG | SNPstream exploratory panel |
| rs1106719-F | GCAGCTACAGTCAGTGGACAT | SNPstream exploratory panel |
| rs1106719-R | ACACCAAGCAGAGCAAGTCT | SNPstream exploratory panel |
| rs1106719-GA-U36 | CACCGCTATCAACAGACTTGTTCCAGGCATTCTAGAATGAGAAAA | SNPstream exploratory panel |
| rs10925022-F | GATGCTTCCTCTGTTCTGGA | SNPstream exploratory panel |
| rs10925022-R | CTGGTTGCTGCTGAGGAC | SNPstream exploratory panel |
| rs10925022-GA-U37 | ACCGCACTAAGCAATGTATCCTCTGTTCTGGAGCTCTCTGGTCAG | SNPstream exploratory panel |
| rs3738447-F | GCTTCCAGGTGCTTCCTC | SNPstream exploratory panel |
| rs3738447-R | ATACATCAATCTTTGCTCATGAAG | SNPstream exploratory panel |
| rs3738447-GA-U38 | ACGTAAGACCACTCAAGACCGCTTCCTCGCCAGCTTCTTCTTGGC | SNPstream exploratory panel |
| rs12137901-F | AATAGAAGGAGAAGGCCAGG | SNPstream exploratory panel |

| | | |
|---|---|---|
| rs12137901-R | TTCATCTCACTTCAATCCACTG | SNPstream exploratory panel |
| rs12137901-GA-U39 | CACGACAAGACAACAGATACGGTGAGTAAGTGTGTTGATAACAGA | SNPstream exploratory panel |
| rs6673762_F | TGGCACTTTGTTACAGCAGA | SNPstream exploratory panel |
| rs6673762_R | TCATCATGTTGTCCAGGCTA | SNPstream exploratory panel |
| rs6673762-GA-U40 | AAGTACCACGTCAACGTCACAGCATTTTGGGAGGCCGAGGCGGGC | SNPstream exploratory panel |
| rs10925014-F | GGCATGGATTTTTACTGCAT | SNPstream exploratory panel |
| rs10925014-R | AGCTCATTGGCTGAACCTG | SNPstream exploratory panel |
| rs10925014-GA-U41 | TACCTATGACCAGCAAGCACACTCCTGGGCTCAAGCTATCCTCCC | SNPstream exploratory panel |
| rs6677787-F | AGAAGGCAGACCATGTGG | SNPstream exploratory panel |
| rs6677787-R | ATCTCTTTTTGCTTTCTCATAAAGG | SNPstream exploratory panel |
| rs6677787-GA-U42 | CTCACTATCTGACAAGCCACTGGGGAGGAGAAGGCGTGGGCCATG | SNPstream exploratory panel |
| rs12239046-F | AGTCGTGAAATCCCCAGAC | SNPstream exploratory panel |
| rs12239046-R | TCCTCTTTGGAAGCCTTCTT | SNPstream exploratory panel |
| rs12239046-GA-U44 | CAAGACCGCAACTAGATACAGGGAGGGGGAAGCTATAACATTAGC | SNPstream exploratory panel |
| rs4378247-F | GAGCACCTCCAATGACAGAG | SNPstream exploratory panel |
| rs4378247-R | GAACCACATTGGGCACTTAG | SNPstream exploratory panel |
| rs4378247-GA-U43 | AGCCGAACTACCACTGAGTACACGGCTCCCAATAAAAATGTGGGG | SNPstream exploratory panel |
| rs4925546-F | GTATCACCAAATAAGGTCACAATCT | SNPstream exploratory panel |
| rs4925546-R | TTTATATTATGGTTCTAGTTTCAATAGATAGG | SNPstream exploratory panel |
| rs4925546-GA-U45 | CTAACTAAGCTACGCCGACACTAGGACTTCCACATATAAAGTCTT | SNPstream exploratory panel |
| rs12128790_F | CCTGATAAGGGGTAAACACTCA | SNPstream exploratory panel |
| rs12128790_R | TTTGCATTTCCCTGATGACT | SNPstream exploratory panel |
| rs12128790-GA-U46 | ACAGATCACTCACCGACTAAACTCTTACAACTCAAAACTAAGAGA | SNPstream exploratory panel |
| rs6700719-F | AAGCTATACAGTGGTGAAATCCTAGT | SNPstream exploratory panel |

| | | |
|---|---|---|
| rs6700719-R | ATTTCAACAGAATCAGATAACTGGA | SNPstream exploratory panel |
| rs6700719-GA-U47 | ACAAGAACTCCATGACTCAACATGGCAGTGATTTCAGTGTGAAAC | SNPstream exploratory panel |
| rs4356092-F | AAATGTTTTCATTTGTTCCACAT | SNPstream exploratory panel |
| rs4356092-R | AGTTTTTTCAGCCTTAAGAAATTATG | SNPstream exploratory panel |
| rs4356092-GA-U48 | CGATCACCTCACTAGAACAAAGAACTTTTAATCTCACCCAGGGGA | SNPstream exploratory panel |
| rs10157379_F | ACGTTGGATGTGCTGGTCATCTGTTCATGC | Sequenom exploratory panel |
| rs10157379_R | ACGTTGGATGTGCCAGAGTCCTTGGCAATG | Sequenom exploratory panel |
| rs10157379_F_CT | GTATTTCTATTTTTCCTGCCT | Sequenom exploratory panel |
| rs10399895_F | ACGTTGGATGTACAAGTGCAGGTTTCTGAAAAG | Sequenom exploratory panel |
| rs10399895_R | ACGTTGGATGACGCAGAGTGGAGGAATAGATA | Sequenom exploratory panel |
| rs10399895_F_CT | TCGTACCCAACAGGTAAT | Sequenom exploratory panel |
| rs10732301_F | ACGTTGGATGCTTAATACGTGGCCATTGCT | Sequenom exploratory panel |
| rs10732301_R | ACGTTGGATGAACCCTGGAGGTGAGTGATT | Sequenom exploratory panel |
| rs10732301_F_AG | ATCCATCCTCCACCTC | Sequenom exploratory panel |
| rs10737805_F | ACGTTGGATGTGTGATCTTCATTTCTGGAGACT | Sequenom exploratory panel |
| rs10737805_R | ACGTTGGATGTTGTCTATCACATGTAAGGGTCTG | Sequenom exploratory panel |
| rs10737805_R_CT | GGGTTGGACGACACAGGCAA | Sequenom exploratory panel |
| rs10754555_F | ACGTTGGATGGAAACAGACAGCAGTGTGCA | Sequenom exploratory panel |
| rs10754555_R | ACGTTGGATGTAAGTACTGACACAATGATAATTGC | Sequenom exploratory panel |
| rs10754555_F_CG | GATGGGAGGTCAGAAACCA | Sequenom exploratory panel |
| rs10754558_F | ACGTTGGATGAGCTTGGGCATCTCCTTTAC | Sequenom exploratory panel |
| rs10754558_R | ACGTTGGATGGAACATCCTCTAACTGAGGC | Sequenom exploratory panel |
| rs10754558_F_CG | ACAATGACAGCATCGGGTGTTGTT | Sequenom exploratory panel |
| rs10802496_F | ACGTTGGATGAGTCAAGCCATGTATGCCTCA | Sequenom exploratory panel |

| | | |
|---|---|---|
| rs10802496_R | ACGTTGGATGGGTTTCACCATATTGGTCAGG | Sequenom exploratory panel |
| rs10802496_F_AT | TACTATTGGCTGGGCG | Sequenom exploratory panel |
| rs10802510_F | ACGTTGGATGATTTGAGGTGATGGATATGCT | Sequenom exploratory panel |
| rs10802510_R | ACGTTGGATGTAAATACTGTAACATTTTTAAAGTTCCA | Sequenom exploratory panel |
| rs10802510_F_CT | ACAATGTATACACATATATCAAAATAT | Sequenom exploratory panel |
| rs10925014_F | ACGTTGGATGATAAAGGCATGGATTTTTACTGC | Sequenom exploratory panel |
| rs10925014_R | ACGTTGGATGCTGTAATCCCAGCTCTCAGG | Sequenom exploratory panel |
| rs10925014_R_CT | AAAATCAGGGAGGCAAGAGG | Sequenom exploratory panel |
| rs10925019_F | ACGTTGGATGCCATGGAGACTGGTTGTTTG | Sequenom exploratory panel |
| rs10925019_R | ACGTTGGATGGCCTCTCCAGAATCATACAG | Sequenom exploratory panel |
| rs10925019_F_CT | TGGTTGTTTGGGACATTTAAC | Sequenom exploratory panel |
| rs10925026_F | ACGTTGGATGGATCACAACCTCAAGATGGG | Sequenom exploratory panel |
| rs10925026_R | ACGTTGGATGTCCATCTAAGAAGCTCCTAC | Sequenom exploratory panel |
| rs10925026_R_GT | CCTCAAGATGGGAATGTTA | Sequenom exploratory panel |
| rs10925039_F | ACGTTGGATGCGCAAGAAGTAGTGCACATT | Sequenom exploratory panel |
| rs10925039_R | ACGTTGGATGAGGGAGGAGAAATTACGTGG | Sequenom exploratory panel |
| rs10925039_F_CT | TGGGCATTTAGCGTAACTATCA | Sequenom exploratory panel |
| rs12048215_F | ACGTTGGATGGGGAAGGAGGAGTACAAATG | Sequenom exploratory panel |
| rs12048215_R | ACGTTGGATGTGGTTTCTGACCTCCCCCG | Sequenom exploratory panel |
| rs12048215_F_AG | GCAGCAGTGTGGGTGTA | Sequenom exploratory panel |
| rs12062001_F | ACGTTGGATGGTGACAGAGCGAGATTCCAT | Sequenom exploratory panel |
| rs12062001_R | ACGTTGGATGACAGAGCAGGGAATGAAACA | Sequenom exploratory panel |
| rs12062001_R_AT | GAGAAAGCAAGCACATTT | Sequenom exploratory panel |
| rs12070953_F | ACGTTGGATGCATGTCCTCATCATCCACAG | Sequenom exploratory panel |

| | | |
|---|---|---|
| rs12070953_R | ACGTTGGATGAATCACCCCCACTCTCAATC | Sequenom exploratory panel |
| rs12070953_F_CT | CAGAGCCCCCTTCATA | Sequenom exploratory panel |
| rs12089794_F | ACGTTGGATGATTCCAGCTTGGGTTTCATC | Sequenom exploratory panel |
| rs12089794_R | ACGTTGGATGAAGATCATGCCACAGCACTC | Sequenom exploratory panel |
| rs12089794_F_AG | TCATCTAGTTCCACCTCA | Sequenom exploratory panel |
| rs1539019_F | ACGTTGGATGCTACCTTCTCTCTGAGTGTC | Sequenom exploratory panel |
| rs1539019_R | ACGTTGGATGGCTATAGGGTAGGGTATCTG | Sequenom exploratory panel |
| rs1539019_F_GT | TCTCTGAGTGTCCTTGGACAAAG | Sequenom exploratory panel |
| rs2027432_F | ACGTTGGATGACCAAGCACTACATCAACCC | Sequenom exploratory panel |
| rs2027432_R | ACGTTGGATGGAGCATTCTCTCTGCAGTTC | Sequenom exploratory panel |
| rs2027432_F_CT | AGTTGGTTCCGATGAC | Sequenom exploratory panel |
| rs4353135_F | ACGTTGGATGGAGAGAATGAAGGAAGCAAG | Sequenom exploratory panel |
| rs4353135_R | ACGTTGGATGTATTACAGTACCTGGCTCAC | Sequenom exploratory panel |
| rs4353135_F_GT | AAGATAGAGGGGTAAATGTAT | Sequenom exploratory panel |
| rs4356092_F | ACGTTGGATGGTCTGGATACACAGCAAATG | Sequenom exploratory panel |
| rs4356092_R | ACGTTGGATGGCCTTAAGAAATTATGTTTCC | Sequenom exploratory panel |
| rs4356092_F_AG | TTAATCTCACCCAGGGGA | Sequenom exploratory panel |
| rs4925546_F | ACGTTGGATGCAATAGATAGGGTGTTATGG | Sequenom exploratory panel |
| rs4925546_R | ACGTTGGATGCCTGTGAAAGAACACAAGTC | Sequenom exploratory panel |
| rs4925546_R_CT | ATGGATTGAATTGTTTCCTC | Sequenom exploratory panel |
| rs4925671_F | ACGTTGGATGCATGAGCTACAATTCATGTG | Sequenom exploratory panel |
| rs4925671_R | ACGTTGGATGGAGTCAAAACCTAGAGAGGG | Sequenom exploratory panel |
| rs4925671_F_CT | CATCATGTGATTAAATGCTGCTA | Sequenom exploratory panel |
| rs6673459_F | ACGTTGGATGCCCCAAGTCTTGGCAATTGA | Sequenom exploratory panel |

| | | |
|---|---|---|
| rs6673459_R | ACGTTGGATGCTTTGCTCTCCTTCCTACTC | Sequenom exploratory panel |
| rs6673459_R_CA | TCAGGAATATTAAATGGTTTAGTTG | Sequenom exploratory panel |
| rs6673762_F | ACGTTGGATGTCATCATGTTGTCCAGGCTAGT | Sequenom exploratory panel |
| rs6673762_R | ACGTTGGATGGTTACAGCAGAAACCACACCA | Sequenom exploratory panel |
| rs6673762_F_CT | AGTCCTGACCTCAGGTGATC | Sequenom exploratory panel |
| rs6700719_F | ACGTTGGATGCCTCAAGCTATACAGTGGTG | Sequenom exploratory panel |
| rs6700719_R | ACGTTGGATGAGAGCCCAGAGTAAAACCAG | Sequenom exploratory panel |
| rs6700719_F_AG | GGAAAGTGATTTCAGTGTGAAAC | Sequenom exploratory panel |
| rs7528887_F | ACGTTGGATGCCTGAAATTTGCTAAGAGAG | Sequenom exploratory panel |
| rs7528887_R | ACGTTGGATGCATGAGTTAACATAGCCACC | Sequenom exploratory panel |
| rs7528887_R_CA | AAGAGATTTTAGATGTTTTTACCAC | Sequenom exploratory panel |
| 9kb_ss107635120_GA-F | ACGTTGGATAGAAAGGGGCAGGGATTAGA | Sequenom Phase II panel |
| 9kb_ss107635120_GA-R | ACGTTGGATCCATTGAGAAACCGTGTGTG | Sequenom Phase II panel |
| 9kb_ss107635120_GA-Pa | ACAAGAGAAACCGTGTGTGTGTG | Sequenom Phase II panel |
| 9kb_ss107635122_CT-F | ACGTTGGATGATACCTTGGCTCTTCCACTC | Sequenom Phase II panel |
| 9kb_ss107635122_CT-R | ACGTTGGATGTCATGACCACACACGCACAG | Sequenom Phase II panel |
| 9kb_ss107635122_CT-P | ACACGCACAGGCAGCCCTGTCACTCACA | Sequenom Phase II panel |
| 9kb_ss107635124_CT-F | ACGTTGGATGTGGGAACTACTGTGTCCATC | Sequenom Phase II panel |
| 9kb_ss107635124_CT-R | ACGTTGGATGGCACTCCTTGACCTTCTTTC | Sequenom Phase II panel |
| 9kb_ss107635124_CT-P | ATCTTTCATCTTCCCACTAC | Sequenom Phase II panel |
| 9kb_ss107635126_GA-F | ACGTTGGATTGAATTCCCGCAGTTTCTTC | Sequenom Phase II panel |
| 9kb_ss107635126_GA-R | ACGTTGGATGCAAATGTCAGGGATAAGAGGT | Sequenom Phase II panel |
| 9kb_ss107635126_GA-Pa | GCGGAATGATATATATCTAAGGGTG | Sequenom Phase II panel |
| 9kb_ss107635130_TA-F | ACGTTGGATGCCTCAACTCCACCTCTTCA | Sequenom Phase II panel |

| | | |
|---|---|---|
| 9kb_ss107635130_TA-R | ACGTTGGATCTATTCCCAGCAGGAGGATG | Sequenom Phase II panel |
| 9kb_ss107635130_TA-Pa | TCTGGCACAGGATGAGAAAGG | Sequenom Phase II panel |
| 9kb_ss107635144_CT-F | ACGTTGGATGCACAAGTGAGGCACAGATAG | Sequenom Phase II panel |
| 9kb_ss107635144_CT-R | ACGTTGGATGAAGCATGAGACCTGAACTGG | Sequenom Phase II panel |
| 9kb_ss107635144_CT-P | CTGAACTGGCTTCAACACCAGGCT | Sequenom Phase II panel |
| rs55646866_CT-F | ACGTTGGATGAAAAAGGCGATGACGTGTGG | Sequenom Phase II panel |
| rs55646866_CT-R | ACGTTGGATGAGTGATGGGCCTATGGGAG | Sequenom Phase II panel |
| rs55646866_CT-P | CTATGGGAGGGAGGATA | Sequenom Phase II panel |
| rs10925032_AG-F | ACGTTGGATGATTCCTCCATTGAAACCCTG | Sequenom Phase II panel |
| rs10925032_AG-R | ACGTTGGATGCATGTTGCAACAGGTGTTCC | Sequenom Phase II panel |
| rs10925032_AG-P | TTCCCTGCTGCCCCAGTTGAT | Sequenom Phase II panel |
| rs10925034_rs10925035-F | ACGTTGGATGGTATCACAGCCTGCGTGTTC | Sequenom Phase II panel |
| rs10925034_rs10925035-R | ACGTTGGATGCGCTCAAGAGTTACCTGCTG | Sequenom Phase II panel |
| rs10925034_AG-P | CTCAAGAGTTACCTGCTGATTGTATC | Sequenom Phase II panel |
| rs10925035_GA-Pa | GGAAACTGCGGGAATTCAG | Sequenom Phase II panel |
| rs10925040_CT-F | ACGTTGGATGAAGCCAAACAACTGAGGAAC | Sequenom Phase II panel |
| rs10925040_CT-R | ACGTTGGATGCTTACTCCATCCTCTGCATC | Sequenom Phase II panel |
| rs10925040_CT-P | ATTTAGCTCCCACTTGTAA | Sequenom Phase II panel |
| rs11583410_CA-F | ACGTTGGATGATGAGAACGGCATAGTGCAG | Sequenom Phase II panel |
| rs11583410_CA-R | ACGTTGGATGATATGCAGTCTTCCACTGGC | Sequenom Phase II panel |
| rs11583410_CA-Pa | ATCCCACCCTGGACCGCTACGTGGCCA | Sequenom Phase II panel |
| rs11802680_GA-F | ACGTTGGATCATCACCGTCAGAGTGTTTGA | Sequenom Phase II panel |
| rs11802680_GA-R | ACGTTGGATGCCCTTCTGCTTGCTTTATG | Sequenom Phase II panel |
| rs11802680_GA-Pa | CGTTATGCTTGCTTTATGTCAC | Sequenom Phase II panel |

| | | |
|---|---|---|
| rs12028142_rs35305980-F | ACGTTGGATGCTCAATTTGCCACAAGTGCC | Sequenom Phase II panel |
| rs12028142_rs35305980-R | ACGTTGGATGACACACCCAGAAGGATGAAG | Sequenom Phase II panel |
| rs12028142_AG-P | AATTGTGAAGGCTTTAGGGGA | Sequenom Phase II panel |
| rs35305980_-A-Pa | GCAACATGAAAAGTGACAACCATAGCT | Sequenom Phase II panel |
| rs12135709_GC-F | ACGTTGGATGCAATGTTGCATCGTTGCAGG | Sequenom Phase II panel |
| rs12135709_GC-R | ACGTTGGATGCAGGTAGCAAGTTCCAAGAG | Sequenom Phase II panel |
| rs12135709_GC-Pa | TCGTTTCTGCCCAGGAC | Sequenom Phase II panel |
| rs12406394_AT-F | ACGTTGGATGGCTGCTCGGCAAAGAAATTG | Sequenom Phase II panel |
| rs12406394_AT-R | ACGTTGGATGCTACGGAAAGATCAGACTGC | Sequenom Phase II panel |
| rs12406394_AT-P | GCTTGAGTGACATTCCA | Sequenom Phase II panel |
| rs12756328_GC-F | ACGTTGGATGTCTCATAATGCTCTCCCTCC | Sequenom Phase II panel |
| rs12756328_GC-R | ACGTTGGATGTTCCTCATGCTGTGAATGAC | Sequenom Phase II panel |
| rs12756328_GC-Pa | ATTATCGGGGGCCCATCAGGGAGTG | Sequenom Phase II panel |
| rs4269805_CT-F | ACGTTGGATCCACTCTTCCTTTCCCAGGT | Sequenom Phase II panel |
| rs4269805_CT-R | ACGTTGGATGCAAAGACACAGAGGCTGCT | Sequenom Phase II panel |
| rs4269805_CT-P | TTCAGAACCAGGAGCCCCAG | Sequenom Phase II panel |
| rs4301663_AG-F | ACGTTGGATCCTCCTTCAATTGCCAAGAC | Sequenom Phase II panel |
| rs4301663_AG-R | ACGTTGGATTCAAACACTCTGACGGTGATG | Sequenom Phase II panel |
| rs4301663_AG-P | GCATTTCTAGAGCGAAGCGTGCT | Sequenom Phase II panel |
| rs4362022_CT-F | ACGTTGGATGGAAGGTGTGCAATGCTTGAG | Sequenom Phase II panel |
| rs4362022_CT-R | ACGTTGGATGTGCTGTTAGGAAGCAACTGG | Sequenom Phase II panel |
| rs4362022_CT-P | CCTTTCGACTGCACCA | Sequenom Phase II panel |
| rs6665526_CT-F | ACGTTGGATGTGCGTGTGTGGTCATGACAG | Sequenom Phase II panel |
| rs6665526_CT-R | ACGTTGGATGGGCTGTATTGATAGGACCAC | Sequenom Phase II panel |

| | | |
|---|---|---|
| rs6665526_CT-P | GCTTGCATCCAGCTCCCGGATTCA | Sequenom Phase II panel |
| rs6672995_AG-F | ACGTTGGATGAGCCCATCAGAGGAAATGTG | Sequenom Phase II panel |
| rs6672995_AG-R | ACGTTGGATGAGTCTGTGTTGGCCTTAGTC | Sequenom Phase II panel |
| rs6672995_AG-P | CCGCGGAGGGACATTTAA | Sequenom Phase II panel |
| rs9988620_CA-F | ACGTTGGATGCCTCAAACCCCTTTGTACTG | Sequenom Phase II panel |
| rs9988620_CA-R | ACGTTGGATGTTTCAGGACAGCCCGGTTTC | Sequenom Phase II panel |
| rs9988620_CA-P | ATGCCCTCACTAATTTAACACTATTC | Sequenom Phase II panel |

**Supplementary Table 2.2: Primers used for the 9kb sequencing experiment**

| ID | Primer Sequence |
|---|---|
| 9kb_set1_F | ATGCACCGTGCTCTCTGTC |
| 9kb_set1_R | GAATCAGGAAATTGGAAGTGAAA |
| 9kb_set2_F | GCAAAGACACAGAGGCTGCT |
| 9kb_set 2_R | CGGCACCTCACAGAAAAAGT |
| 9kb_set3_F | CTGCTCGGCAAAGAAATTGT |
| 9kb_set 3_R | TCCCAGGTGATAGCCTGTTT |
| 9kb_set4_F | GCCACAGCTCACTGTTTTGA |
| 9kb_set 4_R | AAGCTCTTGATTGGCGTGAG |
| 9kb_set5_F | TCTCCTCCCAGATTCTATCAACA |
| 9kb_set5_R | CCATTGAGAAACCGTGTGTG |
| 9kb_set6_F | AGAAAGGGGCAGGGATTAGA |
| 9kb_set6_R | TTCGTGTGAGGGTGAGTGAG |
| 9kb_set7_F | CTTGCTACCTGCCTCTCTGC |
| 9kb_set7_R | GTGCCAGCTAACGTCTCGAT |
| 9kb_set8_F | TTTGCAGTCCTGGTTCAATG |
| 9kb_set8_R | GCAAGACCCCATCTCTATTTATTTT |
| 9kb_set9_F | CCTCTGGATGTCATAATCTGTCAC |
| 9kb_set9_R | CCAGCTCCTTCTCCTCCTCT |
| 9kb_set10_F | AGAGTCCTCAGCCTCGTCAG |
| 9kb_set10_R | TTCCTGGAATATGAAGAGCATTT |
| 9kb_set11_F | ATCCTGCACCTCCACTGC |
| 9kb_set11_R | CACACACACACACACAGAGTTCTT |
| 9kb_set12_F | TGAGGTGGTGTGAGAGCTTG |
| 9kb_set12_R | TCTGTGCCACTCATTGTTCTG |
| 9kb_set13_F | TCTGTGCCTCACTTGTGTTTTC |
| 9kb_set13_R | CATGACTCACCCTGTGTCCA |

**Supplementary Table 2.3: SNPs uncovered while sequencing the 9kb region**

| # | SNP ID | Position (dbSNP129) | Frequency[e] |
|---|---|---|---|
| 1 | rs4925663[a] | 245,681,240 | 0.417 |
| 2 | rs11583410[c] | 245,681,519 | 0.354 |
| 3 | rs12028142[c] | 245,681,875 | 0.25 |
| 4 | rs35305980[c] | 245,681,886 | 0.417 |
| 5 | rs4269805[c] | 245,682,093 | 0.063 |
| 6 | rs6673459[a] | 245,682,135 | 0.071 |
| 7 | rs6698597[a] | 245,682,218 | 0.045 |
| 8 | rs4301663[c] | 245,682,292 | 0.136 |
| 9 | rs10802505 | 245,682,403 | 0.432 |
| 10 | rs11802680[d] | 245,682,407 | 0.333 |
| 11 | rs6665526[c] | 245,682,510 | 0.432 |
| 12 | ss107635122[b,d] | 245,682,563 | 0.024 |
| 13 | rs10925031 | 245,682,666 | 0.024 |
| 14 | rs6674091 | 245,682,723 | 0.024 |
| 15 | rs12406394[c] | 245,682,804 | 0.026 |
| 16 | rs4362022[d] | 245,682,942 | 0.043 |
| 17 | rs4372298 | 245,683,069 | 0.457 |
| 18 | rs10925032[c] | 245,683,202 | 0.022 |
| 19 | rs4333884 | 245,683,346 | 0.457 |
| 20 | rs4593864 | 245,683,356 | 0.457 |
| 21 | rs4518943 | 245,683,389 | 0.457 |
| 22 | rs4436424 | 245,683,425 | 0.478 |
| 23 | rs55903505[b] | 245,683,462 | 0.435 |
| 24 | rs4353135[a] | 245,683,659 | 0.413 |
| 25 | rs56310736[b] | 245,683,709 | 0.022 |
| 26 | rs4266924[a] | 245,683,757 | 0.457 |
| 27 | rs10802506 | 245,683,999 | 0.478 |
| 28 | rs6677999 | 245,684,020 | 0.478 |
| 29 | ss107635120[b,c] | 245,684,132 | 0.022 |
| 30 | ss107635124[b,d] | 245,684,289 | 0.048 |
| 31 | rs6669625 | 245,684,356 | 0.022 |
| 32 | rs10802507 | 245,684,374 | 0.477 |
| 33 | rs10802508 | 245,684,378 | 0.477 |
| 34 | ss107635146[b] | 245,684,452 | 0.021 |
| 35 | rs12745508 | 245,684,602 | 0.477 |
| 36 | rs12135709[c] | 245,684,694 | 0.406 |
| 37 | rs10925033 | 245,684,862 | 0.438 |
| 38 | rs10925034[d] | 245,684,892 | 0.042 |
| 39 | rs10925035[d] | 245,684,900 | 0.438 |
| 40 | ss107635126[b,d] | 245,684,966 | 0.021 |
| 41 | rs6672845 | 245,685,068 | 0.5 |
| 42 | rs10925036 | 245,685,254 | 0.438 |

| | | | |
|---|---|---|---|
| 43 | rs9700400 | 245,685,366 | 0.438 |
| 44 | ss107635128[b] | 245,685,521 | 0.438 |
| 45 | rs10925038 | 245,685,641 | 0.479 |
| 46 | ss107635130[b,d] | 245,685,647 | 0.042 |
| 47 | rs9988617 | 245,685,669 | 0.479 |
| 48 | rs9988501 | 245,685,739 | 0.479 |
| 49 | rs9287213 | 245,685,750 | 0.042 |
| 50 | rs9988571 | 245,685,778 | 0.479 |
| 51 | rs9988620[c] | 245,685,814 | 0.479 |
| 52 | ss107635132[b] | 245,685,884 | 0.438 |
| 53 | rs9988621 | 245,685,903 | 0.479 |
| 54 | rs9988572 | 245,685,933 | 0.042 |
| 55 | ss107635133[b] | 245,685,965 | 0.438 |
| 56 | ss107635136[b] | 245,686,061 | 0.438 |
| 57 | ss107635138[b] | 245,686,130 | 0.438 |
| 58 | ss107635140[b] | 245,686,136 | 0.438 |
| 59 | rs4925664 | 245,686,610 | 0.043 |
| 60 | rs12756328[d] | 245,686,690 | 0.143 |
| 61 | rs4925666 | 245,686,879 | 0.476 |
| 62 | rs55646866[b,c] | 245,687,008 | 0.457 |
| 63 | rs4925667 | 245,687,277 | 0.5 |
| 64 | rs10732301 | 245,687,457 | 0.043 |
| 65 | ss107635142[b] | 245,687,462 | 0.457 |
| 66 | rs6672995[c] | 245,687,656 | 0.479 |
| 67 | rs4925669 | 245,687,946 | 0.479 |
| 68 | rs11267736 | 245,688,318 | 0.475 |
| 69 | ss107635144[b,d] | 245,688,775 | 0.295 |
| 70 | rs10732302 | 245,688,851 | 0.479 |
| 71 | rs55775744[b] | 245,688,902 | 0.435 |
| 72 | rs10737807 | 245,688,941 | 0.5 |
| 73 | rs34837390 | 245,688,958 | 0.479 |
| 74 | rs10733113[a] | 245,688,980 | 0.479 |
| 75 | rs10925039[a] | 245,689,184 | 0.146 |
| 76 | rs10925040[d] | 245,689,321 | 0.417 |
| 77 | rs10925041 | 245,689,331 | 0.417 |
| 78 | rs10754561[a] | 245,689,423 | 0.083 |
| 79 | rs4925671 | 245,689,497 | 0.438 |

**Supplementary Table 2.3 footnotes:** [a] SNPs (n=8) observed in the sequencing experiment that had already been genotyped in the exploratory phase; [b] Novel SNPs (n=18) uncovered in the sequencing experiment as of dbSNP Build 126; 4 of these SNPs are now part of dbSNP Build 129; [c] SNPs (n=13) that were prioritized for the second phase of the genotyping experiments; [d] SNPs (n=11) observed in the sequencing experiment that were selected as tag for the second phase of the genotyping experiments using a pairwise tagging approach; [e] The minor allele frequency was estimated using the 24 samples sequenced for the 9kb region.

**Supplementary Table 2.4: Association results between *NLRP3* gene expression levels and associated genotypes**

| | Unstimulated Monocytes | | | Unstimulated Peripheral Blood Cells | | |
|---|---|---|---|---|---|---|
| **SNP ID** | **Slope** | **$R^2$** | ***p* value** | **Slope** | **$R^2$** | ***p* value** |
| **rs4353135** | 0.856 ± 0.321 | 0.1968 | 0.0124 | 0.677± 0.203 | 0.2833 | 0.00246 |
| **rs4266924** | 0.356 ± 0.578 | 0.01292 | 0.543 | 0.753 ± 0.450 | 0.09112 | 0.105 |
| **rs55646866** | 0.356 ± 0.578 | 0.01292 | 0.543 | 0.800 ± 0.442 | 0.1080 | 0.0817 |
| **rs6672995** | -0.011 ± 0.466 | 0.0000207 | 0.980 | 0.301 ± 0.357 | 0.02577 | 0.405 |
| **ss107635144** | -0.134 ± 0.491 | 0.002545 | 0.787 | 0.603 ± 0.415 | 0.07025 | 0.157 |
| **rs10733113** | -0.134 ± 0.491 | 0.002545 | 0.787 | 0.603 ± 0.415 | 0.07025 | 0.157 |

**Supplementary Table 2.5: Association results between IL-1$\beta$ levels and associated genotypes**

| | Unstimulated | | | LPS 1µg/ml | | |
|---|---|---|---|---|---|---|
| **SNP ID** | **Slope** | **$R^2$** | ***p* value** | **Slope** | **$R^2$** | ***p* value** |
| **rs4353135** | -0.0916 ± 0.102 | 0.02786 | 0.378 | -0.623 ± 2.85 | 0.001767 | 0.829 |
| **rs4266924** | 0.0045 ± 0.166 | 0.0000260 | 0.979 | 4.44 ± 4.485 | 0.0350 | 0.331 |
| **rs55646866** | 0.0568 ± 0.178 | 0.00363 | 0.752 | 6.51 ± 4.733 | 0.06554 | 0.180 |
| **rs6672995** | 0.2610 ± 0.127 | 0.1346 | 0.0502 | 9.85 ± 3.285 | 0.2569 | 0.00591 |
| **ss107635144** | 0.0784 ± 0.149 | 0.00971 | 0.604 | 4.62 ± 4.041 | 0.04618 | 0.263 |
| **rs10733113** | 0.0784 ± 0.149 | 0.00971 | 0.604 | 4.62 ± 4.041 | 0.04618 | 0.263 |

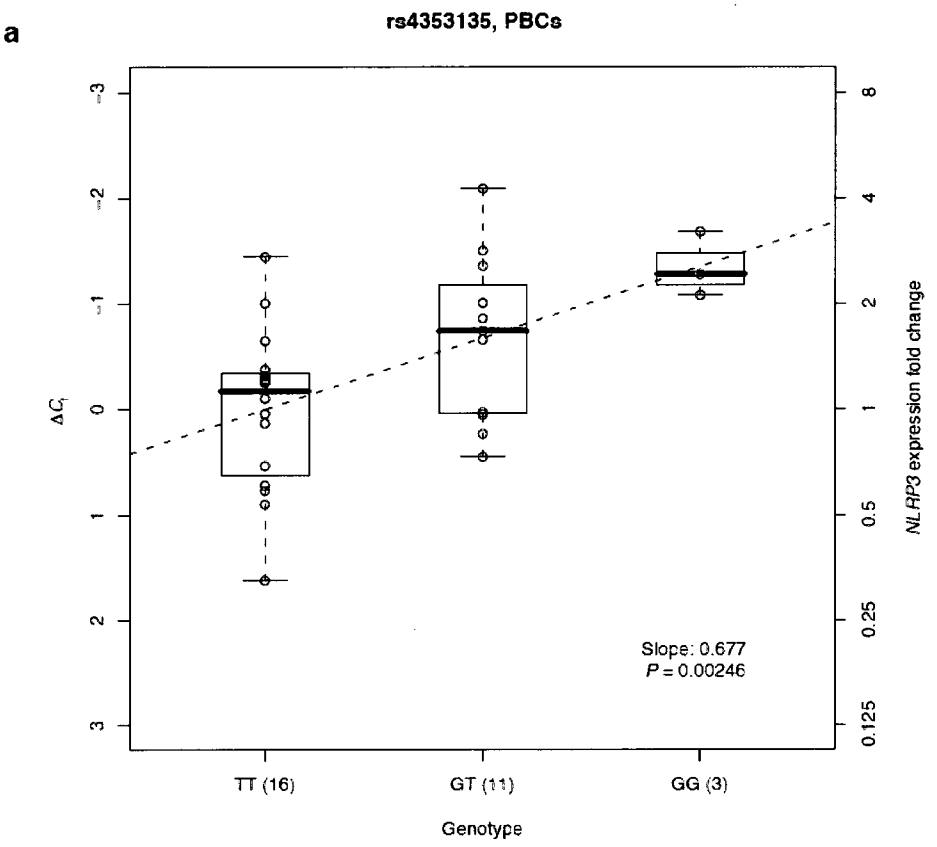
**Supplementary Table 2.6: Linkage disequilibrium between SNPs from Table 2.1 and SNPs genotyped in the *NLRP3* region in recently published genome-wide association studies (GWAS) for CD**

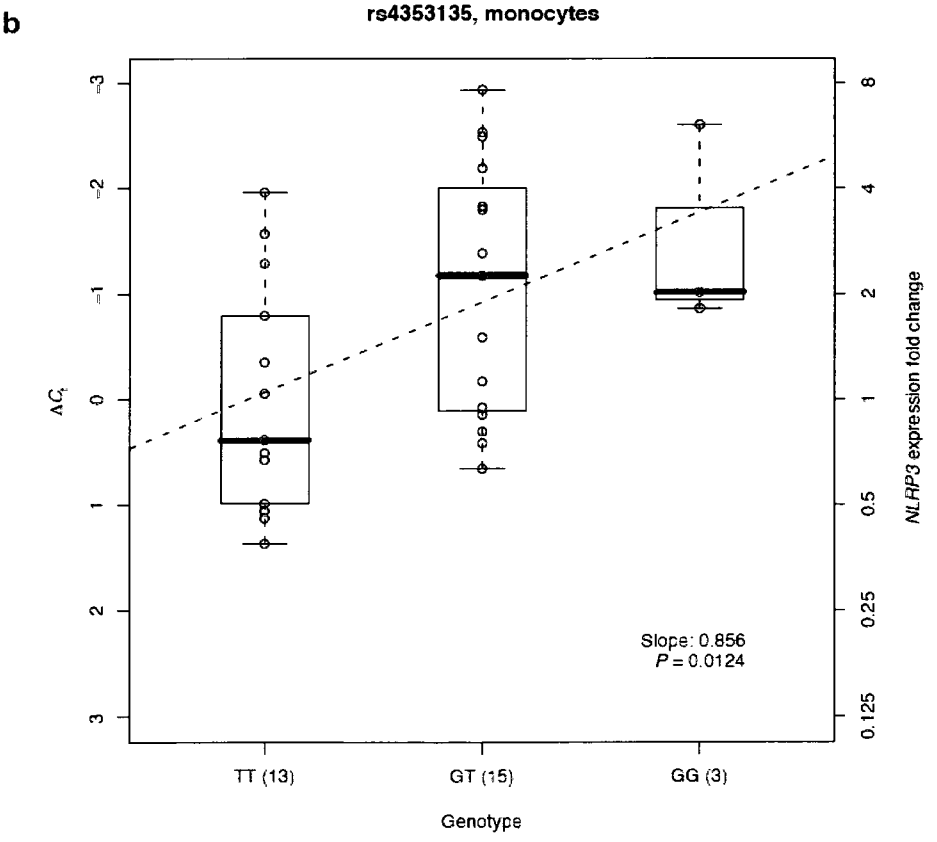
| | | *NLRP3* replicated SNPs[a, b] | | | |
|---|---|---|---|---|---|
| | | rs4353135 | rs4266924 | rs6672995 | rs10733113 |
| GWAS[a] | rs7529058 | 0.015 | 0.006 | 0 | 0.027 |
| | rs3738448 | 0.052 | 0.023 | 0.003 | 0.010 |
| | rs4612666 | 0.020 | 0.033 | 0 | 0.009 |
| | rs12143966 | 0.048 | 0.027 | 0.103 | 0.070 |
| | rs4925659 | 0.034 | 0.014 | 0.078 | 0.049 |
| | rs12065526 | 0.035 | 0.002 | 0.163 | 0.068 |

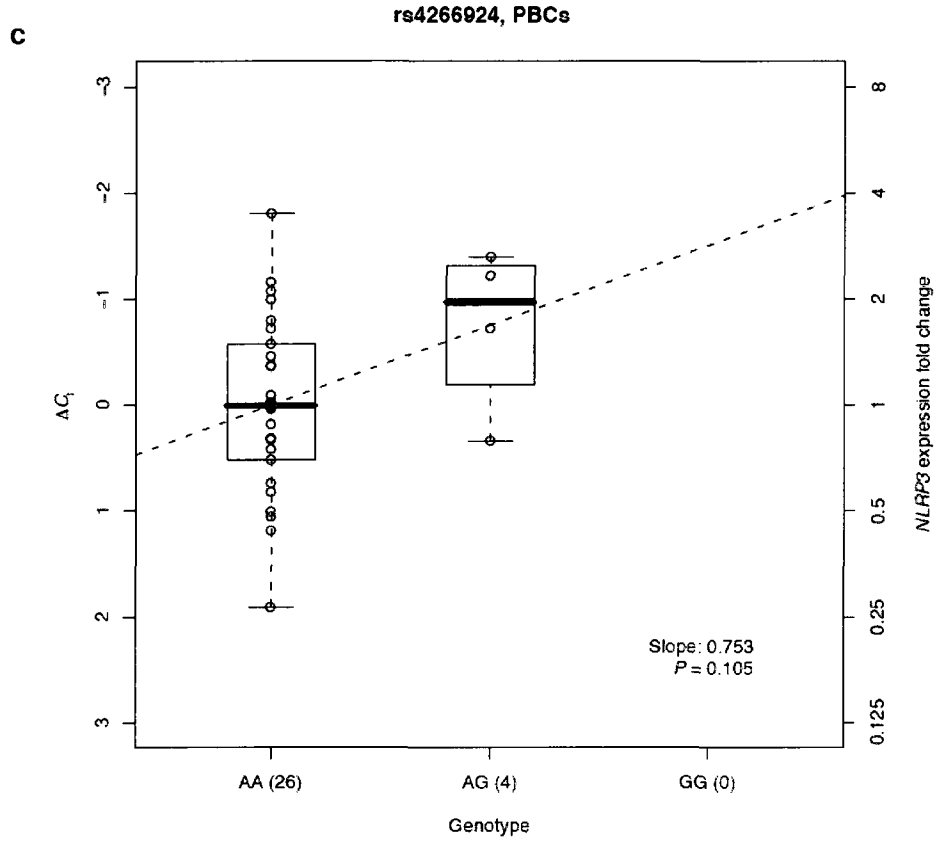
[a] $r^2$ measure of linkage disequilibrium between the 4 replicated SNPs and SNPs present on the Affymetrix 500k and HumanHap300 arrays. SNPs are restricted to the region shown in **Figure 2.1** and that were not genotyped in the present study; [b] rs55646866 and ss107635144 were excluded from this table because these 2 SNPs were uncovered as part of the resequencing effort in this project. There are thus no genotypes available from the HapMap project for these 2 SNPs.

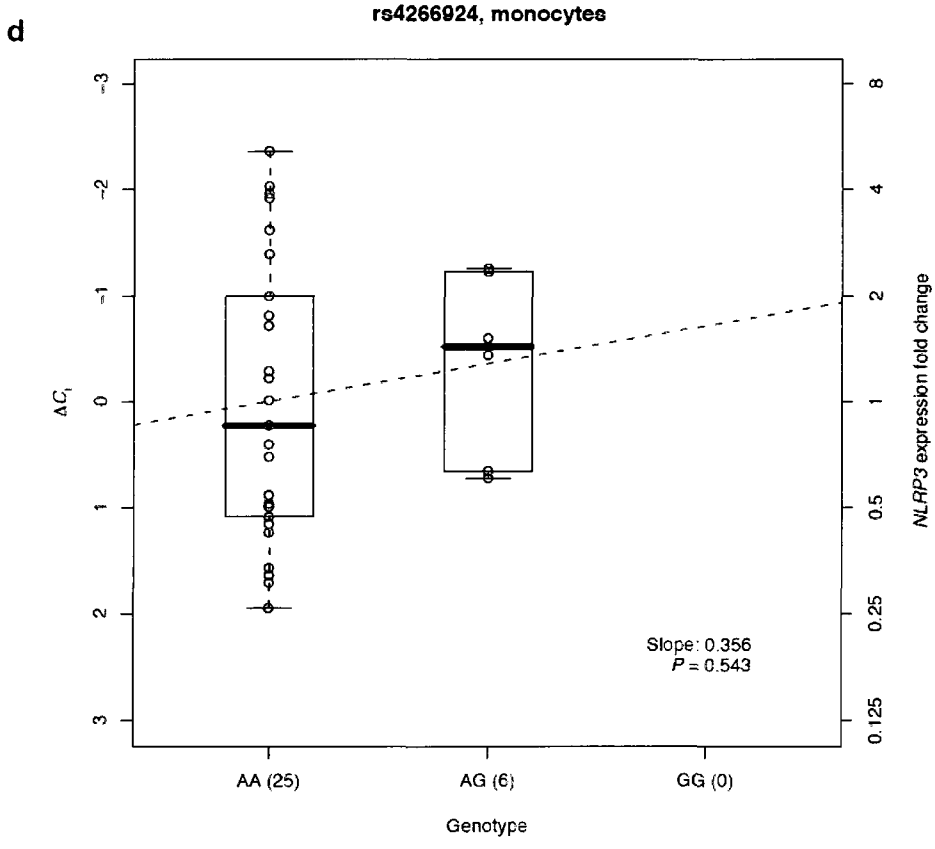
### 2.2.11 Supplementary Figures

**Supplementary Figure 2.1: Association results between *NLRP3* gene expression levels and associated genotypes.** Linear regression analysis of *NLRP3* mRNA level versus the genotypes of the six SNPs from **Table 2.1** (rs4353135, rs4266924, rs55646866, rs6672995, ss107635144, and rs10733113) in DNA-RNA matched freshly isolated peripheral blood cells (a,c,e,g,i,k; PBCs; $n$ = 30) and monocytes (b,d,f,h,j,l; $n$ = 31) obtained from healthy individuals. Genotypes of the six Crohn's disease (CD)-associated SNPs (**Table 2.1**) were obtained by sequencing. Mean threshold cycle ($C_t$) was calculated for each sample from three replicates and then used to calculate relative expression levels ($\Delta C_t$), which is the difference between *NLRP3* $C_t$ and endogenous control 18S RNA $C_t$. Fold change in *NLRP3* expression was calculated using the ABI's comparative $C_t$ method (see Methods), using as a reference the average $\Delta C_t$ of the homozygous for the risk allele of each SNP. $\Delta C_t$ (a-l) for each individual is shown in red; regression lines are shown as dashed lines (a-l).

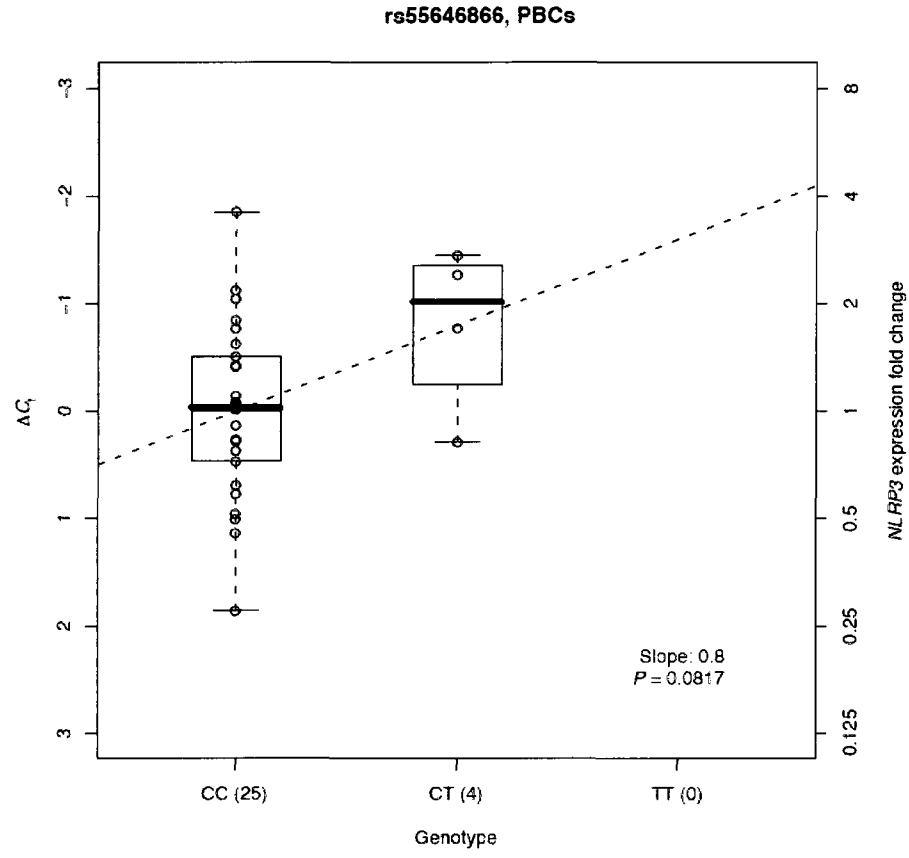

a
rs4353135, PBCs
ΔCt
NLRP3 expression fold change
Slope: 0.677
P = 0.00246
TT (16)
GT (11)
GG (3)
Genotype


b
rs4353135, monocytes
ΔCt
NLRP3 expression fold change
Slope: 0.856
P = 0.0124
TT (13)
GT (15)
GG (3)
Genotype

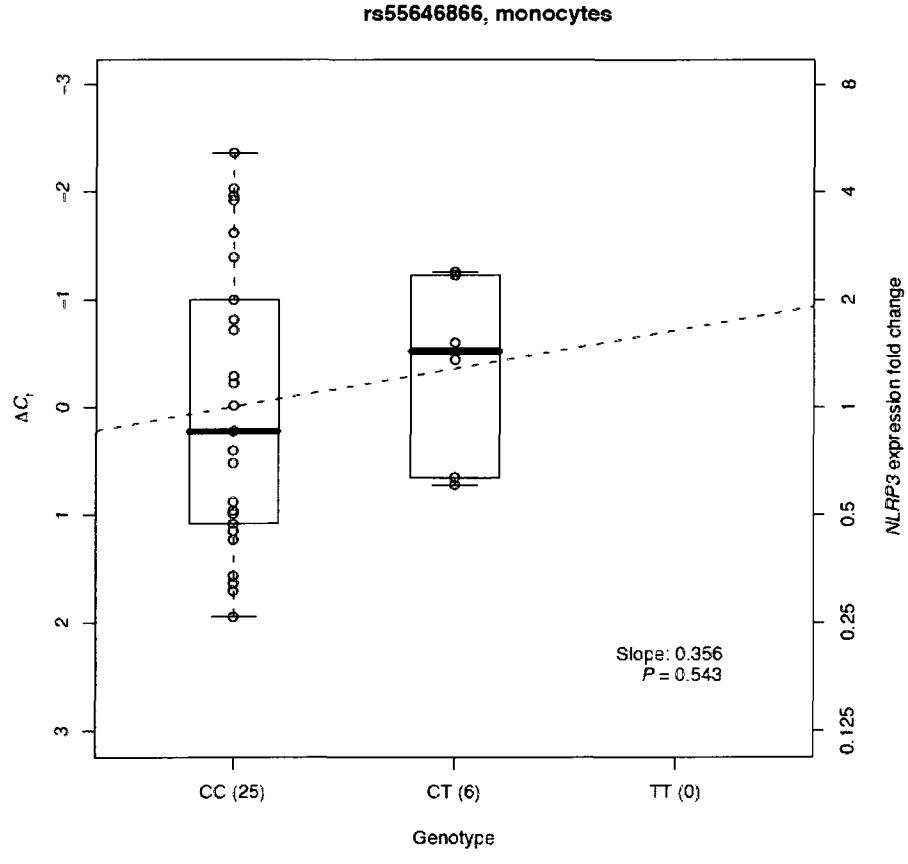

c
rs4266924, PBCs
ΔCt
NLRP3 expression fold change
Slope: 0.753
P = 0.105
AA (26)
AG (4)
GG (0)
Genotype


d
rs4266924, monocytes
ΔCt
NLRP3 expression fold change
Slope: 0.356
P = 0.543
AA (25)
AG (6)
GG (0)
Genotype

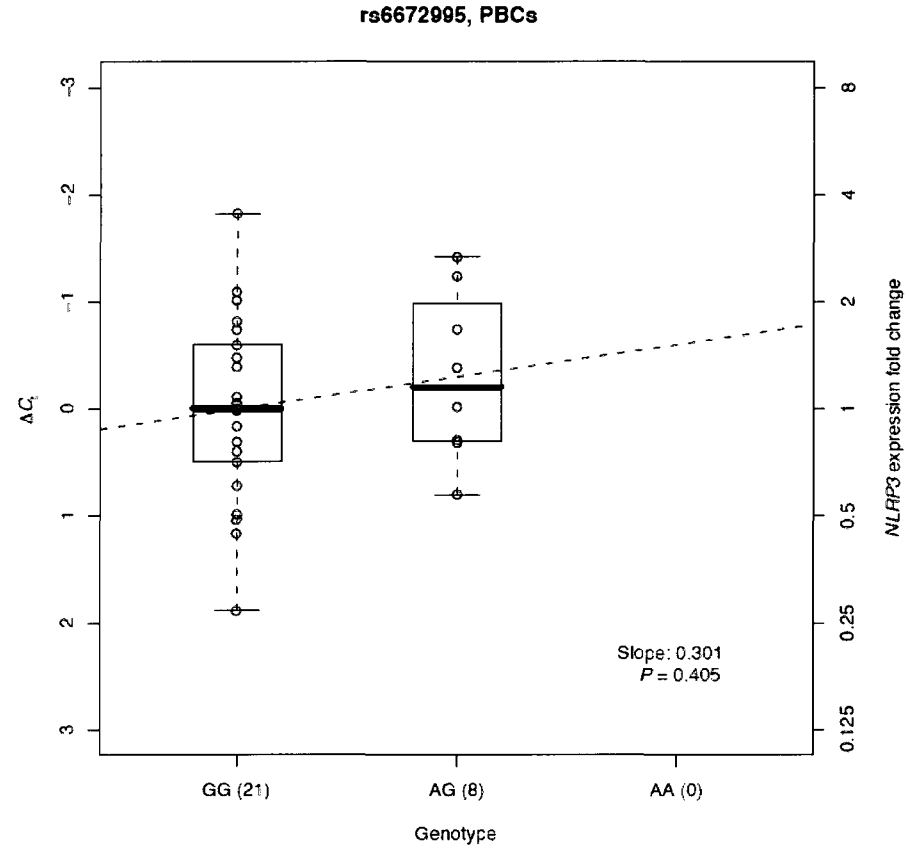
e


rs55646866, PBCs
ΔCt
NLRP3 expression fold change
Slope: 0.8
P = 0.0817
CC (25)
CT (4)
TT (0)
Genotype

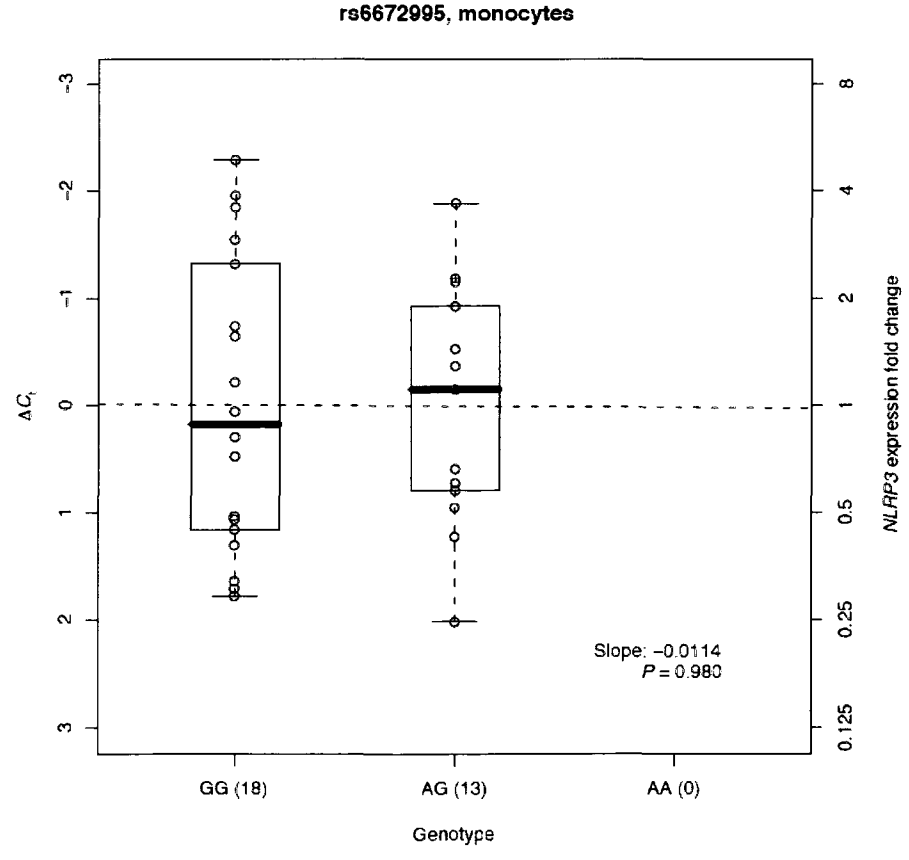
f


rs55646866, monocytes
ΔCt
NLRP3 expression fold change
Slope: 0.356
P = 0.543
CC (25)
CT (6)
TT (0)
Genotype

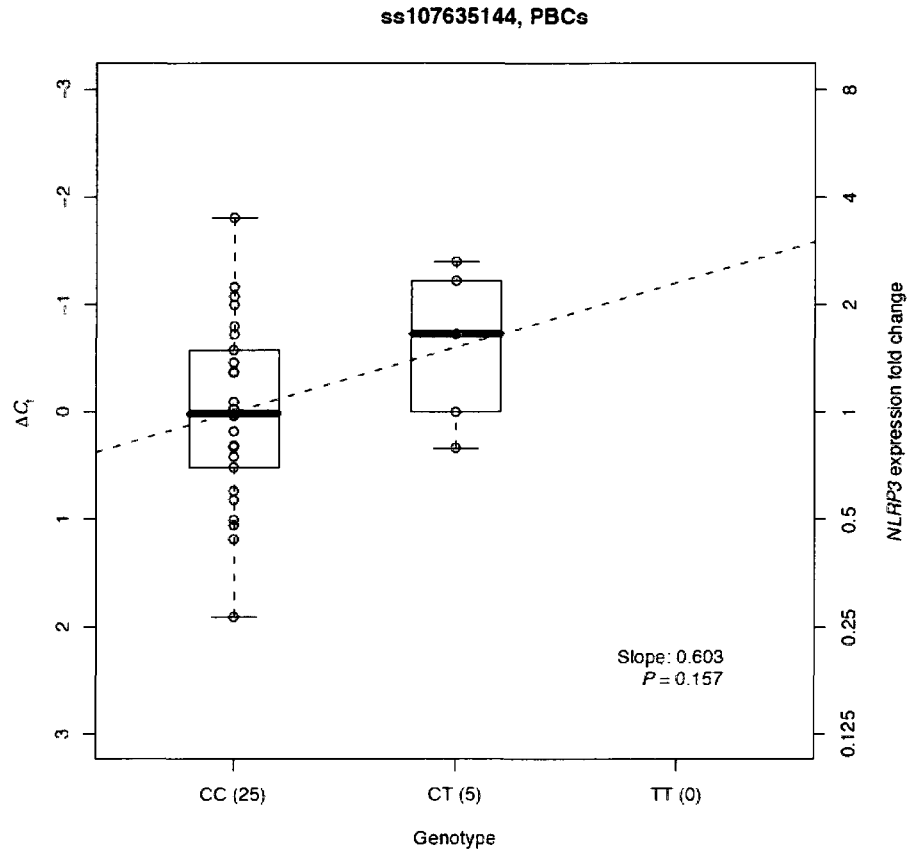
g


rs6672995, PBCs
ΔCt
NLRP3 expression fold change
Slope: 0.301
P = 0.405
GG (21)
AG (8)
AA (0)
Genotype

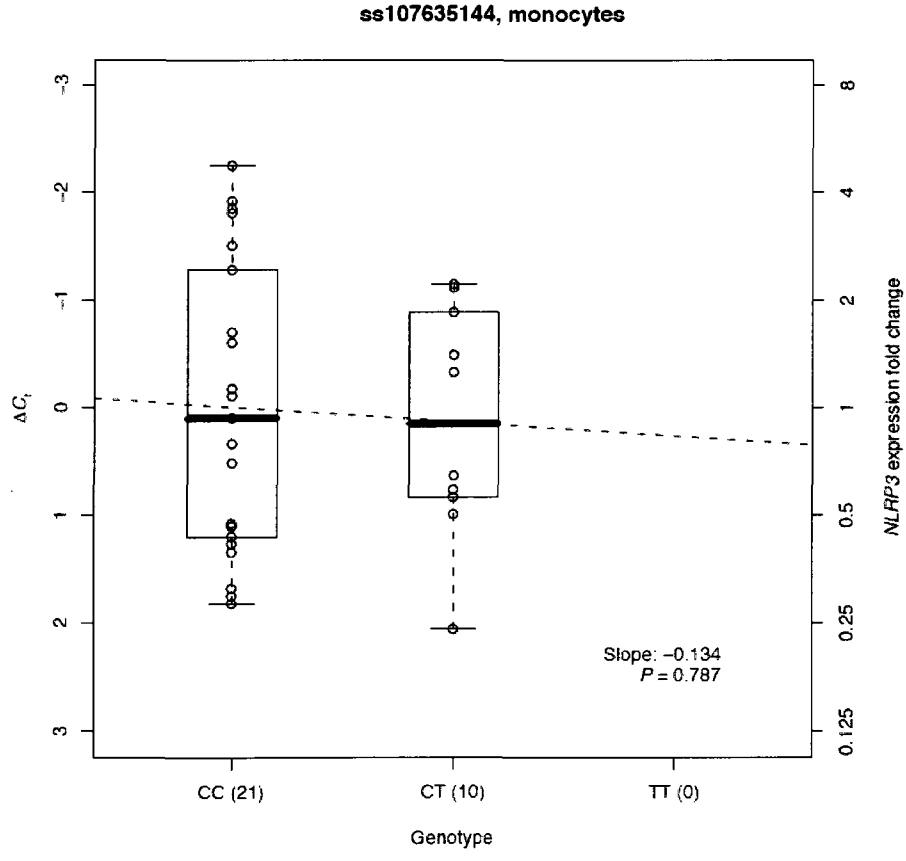
h


rs6672995, monocytes
ΔCt
NLRP3 expression fold change
Slope: −0.0114
P = 0.980
GG (18)
AG (13)
AA (0)
Genotype

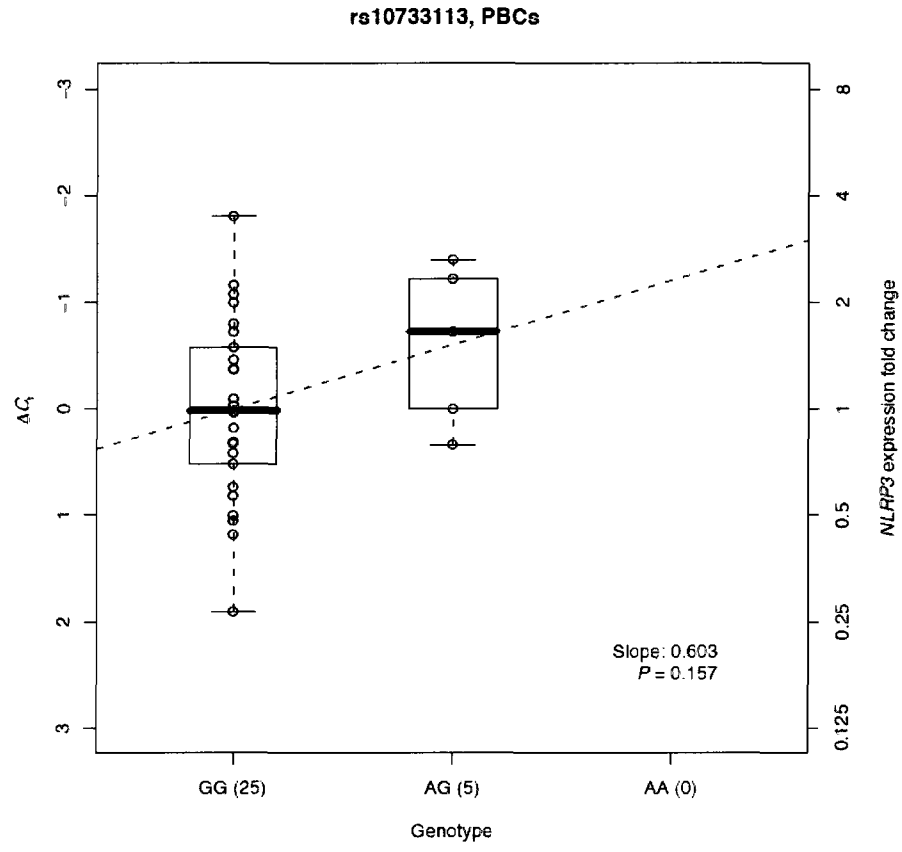
i


ss107635144, PBCs
ΔCt
NLRP3 expression fold change
Slope: 0.603
P = 0.157
CC (25)
CT (5)
TT (0)
Genotype

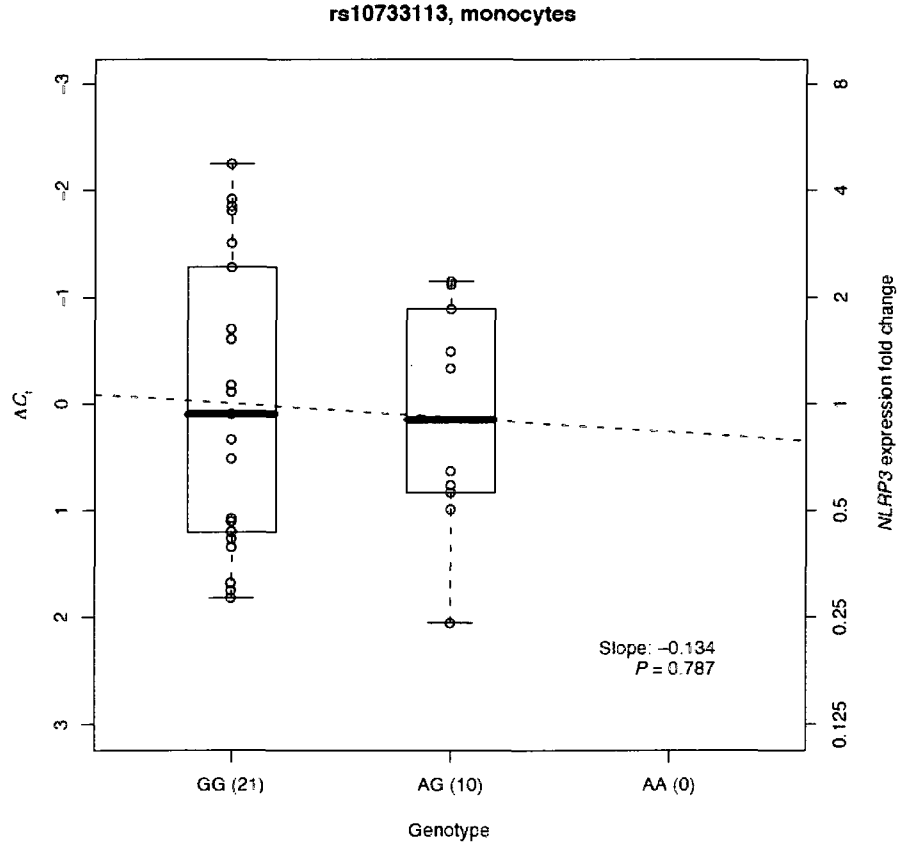
j


ss107635144, monocytes
ΔCt
NLRP3 expression fold change
Slope: −0.134
P = 0.787
CC (21)
CT (10)
TT (0)
Genotype

k

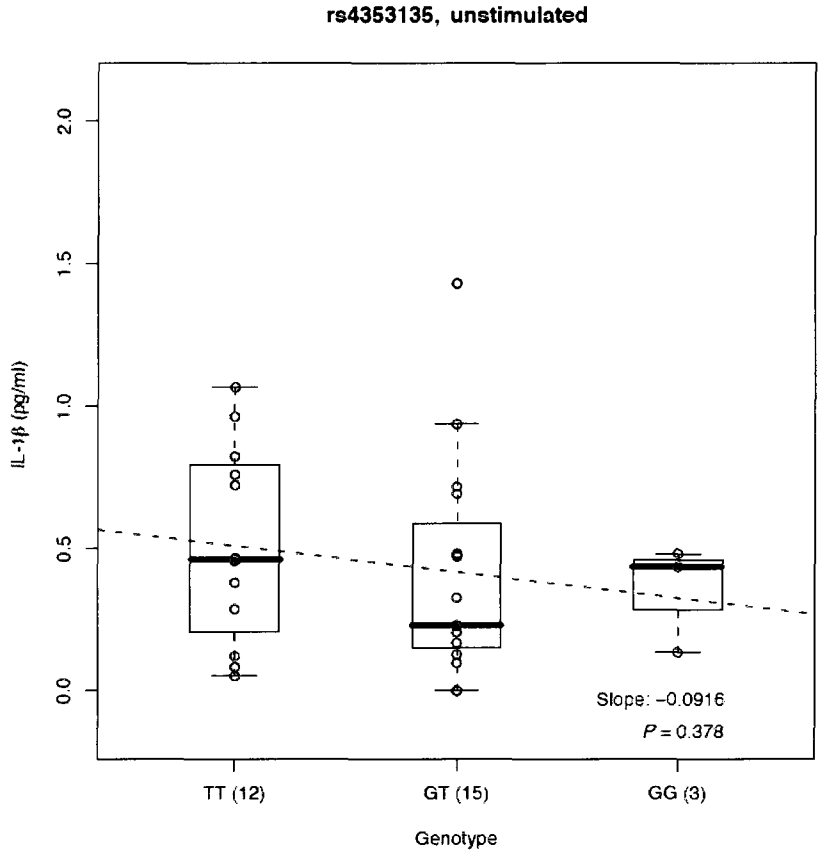

rs10733113, PBCs
ΔCt
NLRP3 expression fold change
Slope: 0.603
P = 0.157
GG (25)
AG (5)
AA (0)
Genotype

l

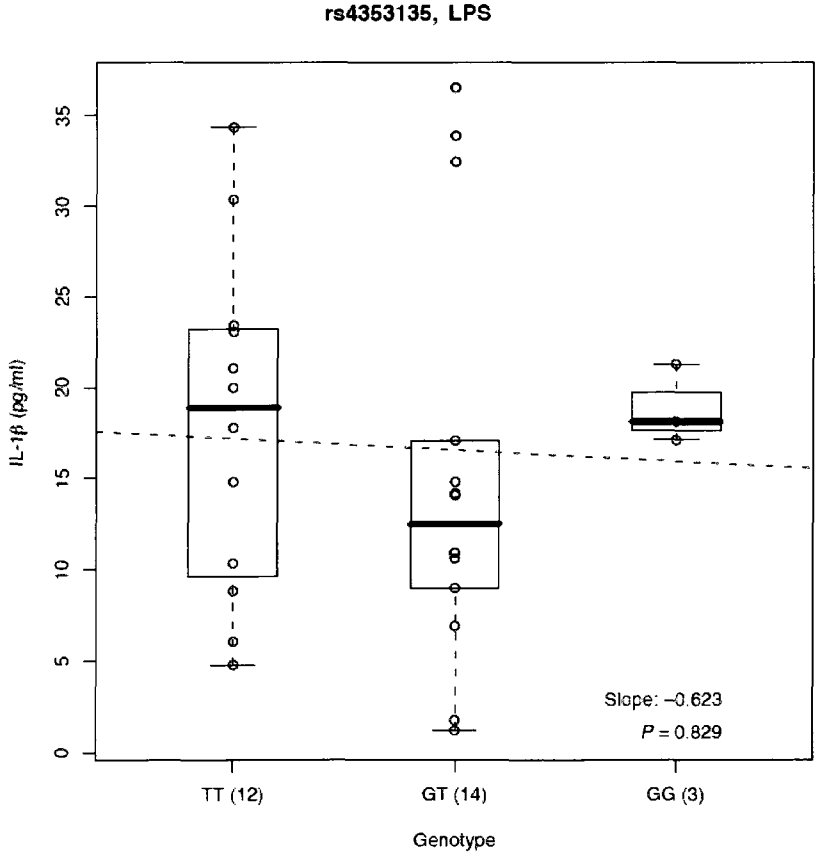

rs10733113, monocytes
ΔCt
NLRP3 expression fold change
Slope: –0.134
P = 0.787
GG (21)
AG (10)
AA (0)
Genotype

**Supplementary Figure 2.2: Association results between IL-1$\beta$ levels and associated genotypes.** Linear regression analysis of IL-1β production (ρg/ml) versus the genotypes of the six SNPs from **Table 2.1** (rs4353135, rs4266924, rs55646866, rs6672995, ss107635144 and rs10733113), obtained from sequencing, for unstimulated (a,c,e,g,i,k) and the LPS-stimulated (b,d,f,h,j,l; 1.0 μg/ml) conditions after 3 h of incubation. The IL-1$\beta$ level (a-l) for each individual is shown in red; regression lines are shown as dashed lines (a-l).

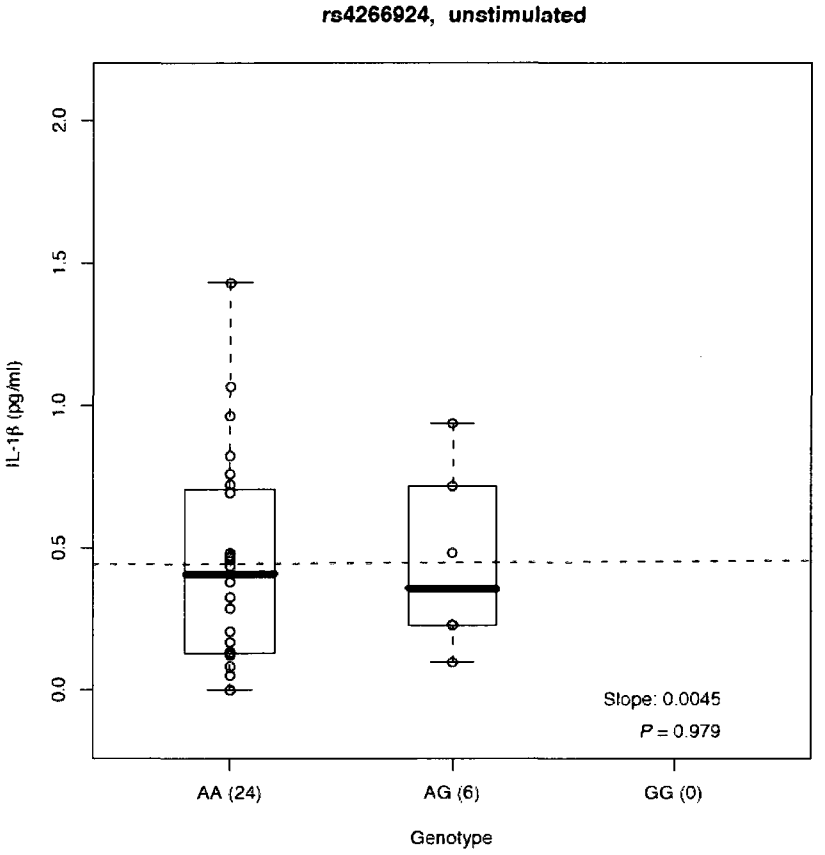
**a**


rs4353135, unstimulated
IL-1β (pg/ml)
0.0
0.5
1.0
1.5
2.0
Slope: −0.0916
P = 0.378
TT (12)
GT (15)
GG (3)
Genotype

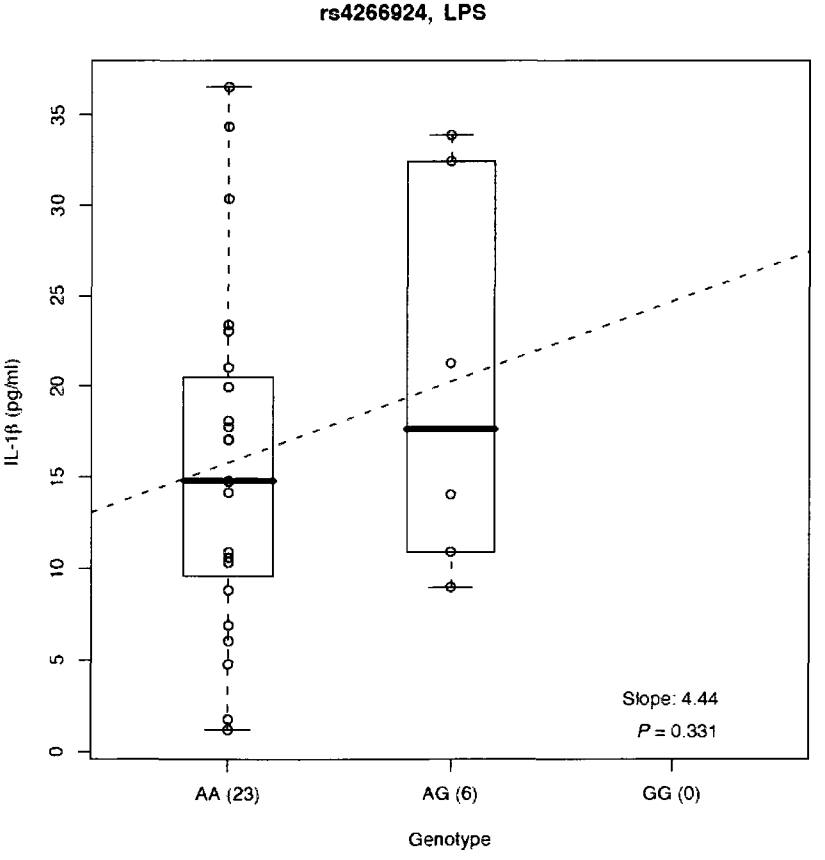
**b**


rs4353135, LPS
IL-1β (pg/ml)
0
5
10
15
20
25
30
35
Slope: −0.623
P = 0.829
TT (12)
GT (14)
GG (3)
Genotype

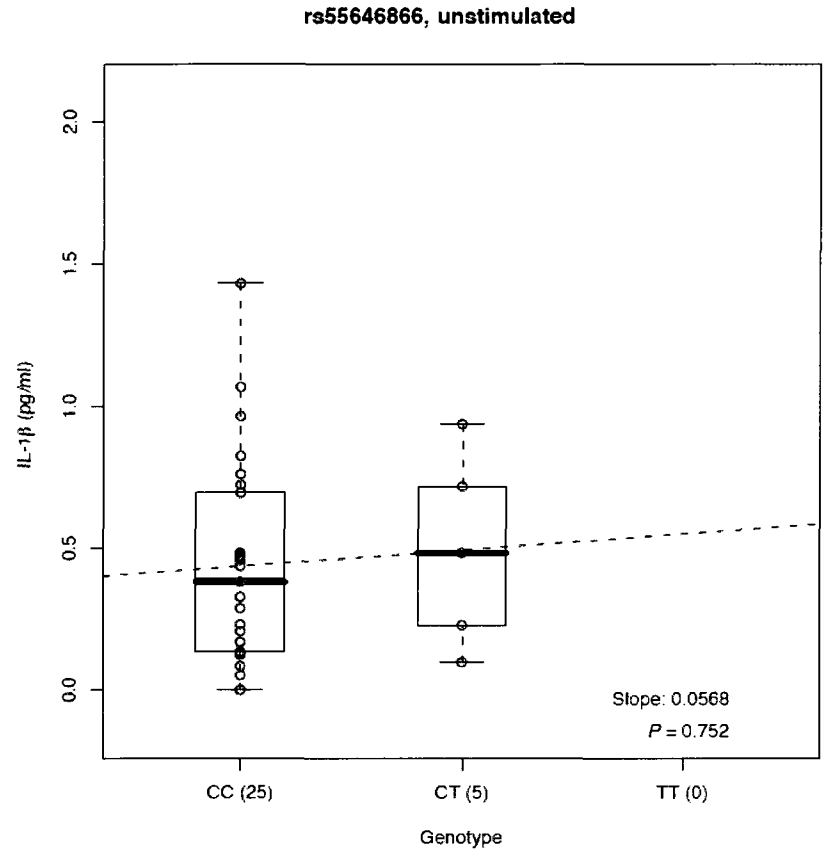
c


rs4266924, unstimulated
IL-1β (pg/ml)
0.0
0.5
1.0
1.5
2.0
Slope: 0.0045
P = 0.979
AA (24)
AG (6)
GG (0)
Genotype

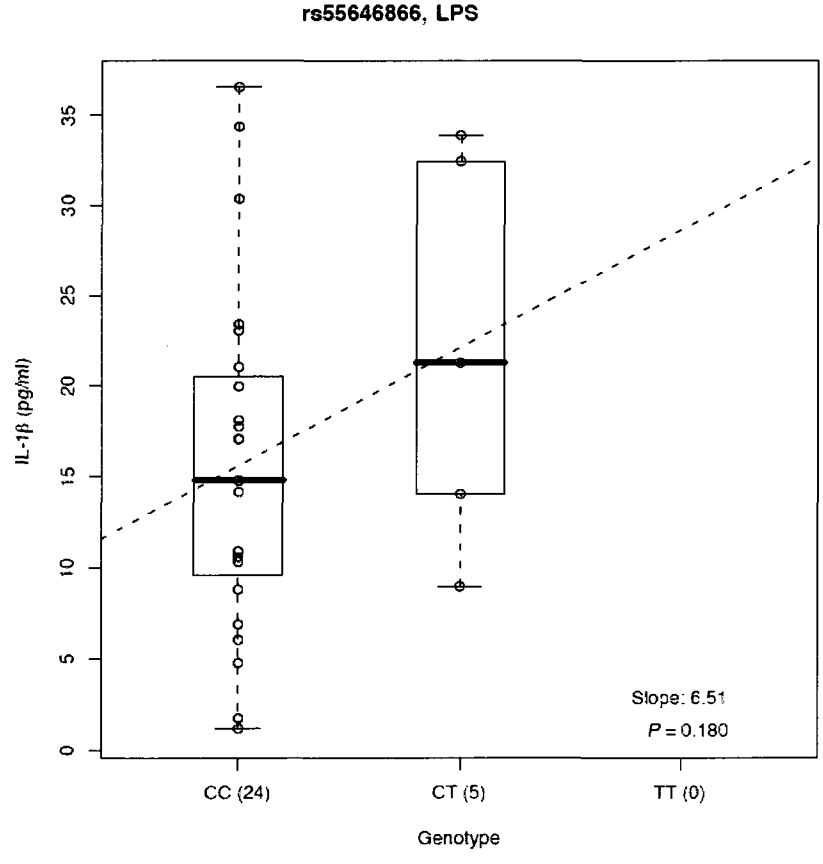
d


rs4266924, LPS
IL-1β (pg/ml)
0
5
10
15
20
25
30
35
Slope: 4.44
P = 0.331
AA (23)
AG (6)
GG (0)
Genotype

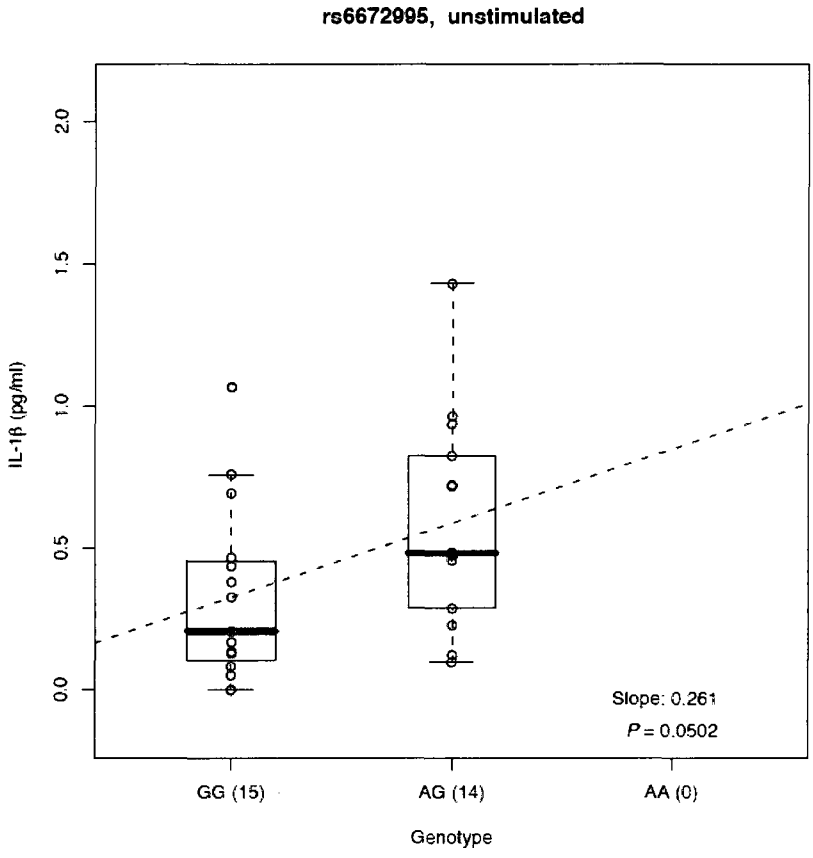
e


rs55646866, unstimulated
IL-1β (pg/ml)
0.0
0.5
1.0
1.5
2.0
Slope: 0.0568
P = 0.752
CC (25)
CT (5)
TT (0)
Genotype

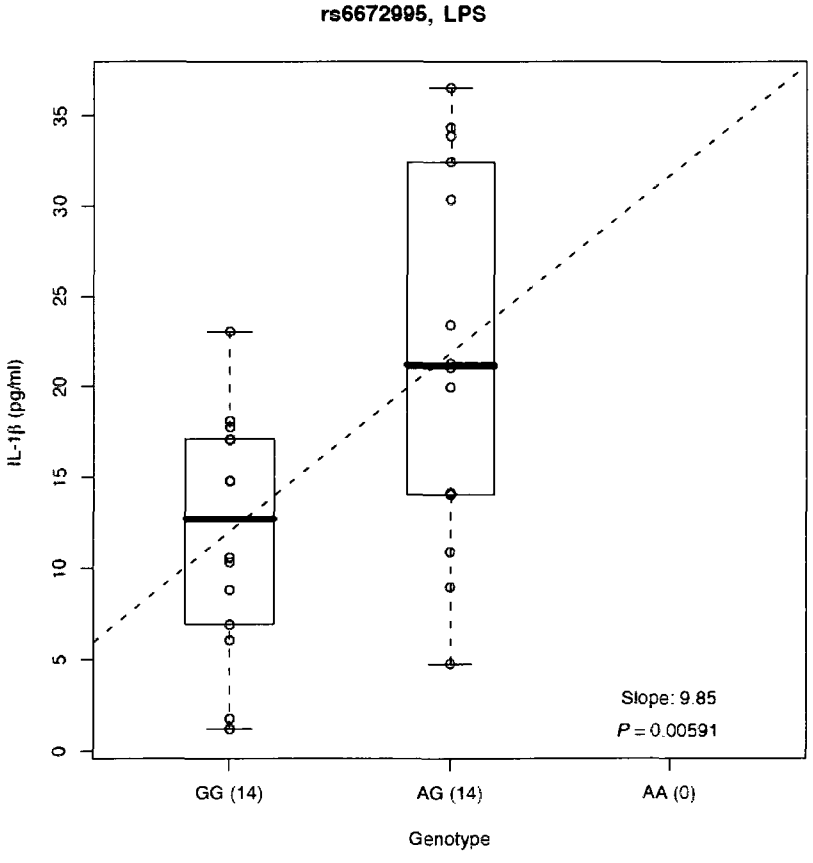
f


rs55646866, LPS
IL-1β (pg/ml)
0
5
10
15
20
25
30
35
Slope: 6.51
P = 0.180
CC (24)
CT (5)
TT (0)
Genotype

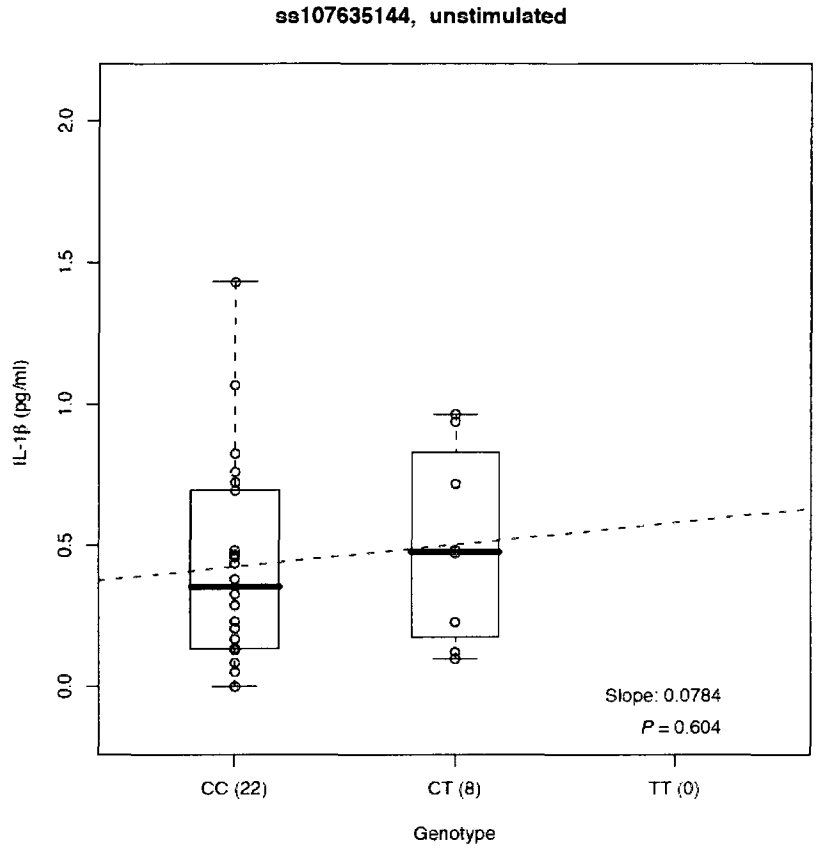
g


rs6672995, unstimulated
IL-1β (pg/ml)
0.0
0.5
1.0
1.5
2.0
Slope: 0.261
P = 0.0502
GG (15)
AG (14)
AA (0)
Genotype

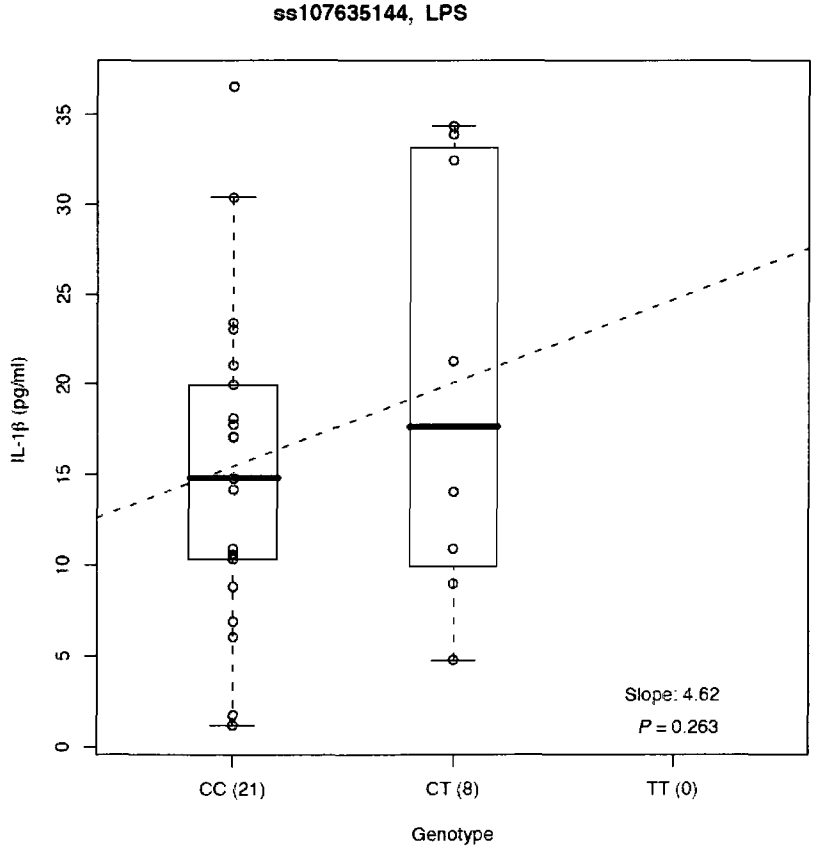
h


rs6672995, LPS
IL-1β (pg/ml)
0
5
10
15
20
25
30
35
Slope: 9.85
P = 0.00591
GG (14)
AG (14)
AA (0)
Genotype

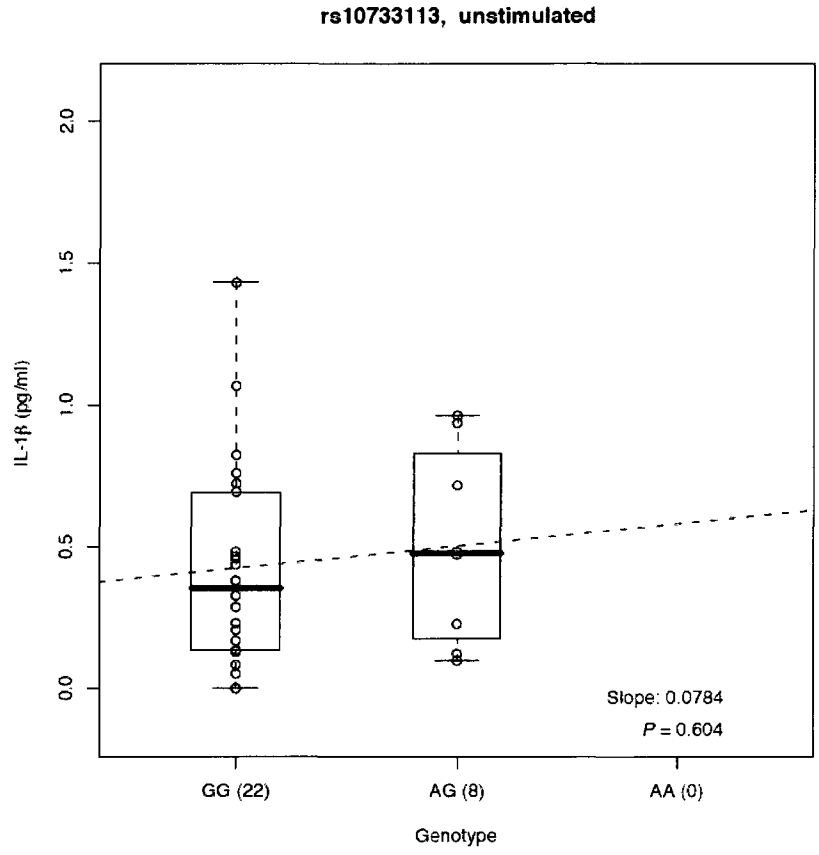
i


ss107635144, unstimulated
IL-1β (pg/ml)
0.0
0.5
1.0
1.5
2.0
Slope: 0.0784
P = 0.604
CC (22)
CT (8)
TT (0)
Genotype

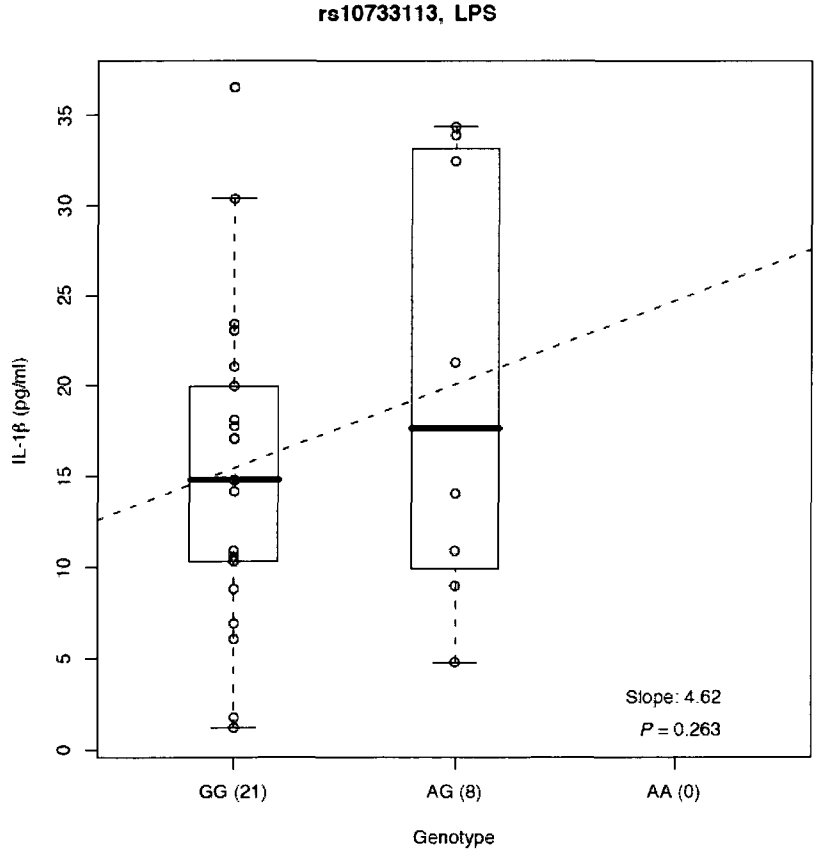
j


ss107635144, LPS
IL-1β (pg/ml)
0
5
10
15
20
25
30
35
Slope: 4.62
P = 0.263
CC (21)
CT (8)
TT (0)
Genotype

k


rs10733113, unstimulated
IL-1β (pg/ml)
0.0
0.5
1.0
1.5
2.0
Slope: 0.0784
P = 0.604
GG (22)
AG (8)
AA (0)
Genotype

l


rs10733113, LPS
IL-1β (pg/ml)
0
5
10
15
20
25
30
35
Slope: 4.62
P = 0.263
GG (21)
AG (8)
AA (0)
Genotype

# 3 INVESTIGATION OF THE *MEFV* GENE IN CROHN'S DISEASE AND ULCERATIVE COLITIS

The manuscript has been submitted to *PLoS ONE*

**Reference:**

Villani AC, Lemire M, Louis E, Silverberg MS, Collette C, Nimmo ER, Renaud Y, Brunet S, Fortin G, Belaiche J, Bitton A, Gaudet D, Cohen A, Rioux JD, Arnott IDR, Rutgeerts P, Satsangi J, Vermeire S, Hudson TJ, Franchimont D. Genetic variation in the familial Mediterranean fever gene (*MEFV*) and risk for inflammatory bowel disease. *PLoS ONE* (submitted).

**Affiliation of co-authors:**

Alexandra-Chloé Villani[1,2], Mathieu Lemire[3], Edouard Louis[4], Mark S. Silverberg[5], Catherine Collette[1], Elaine R. Nimmo[6], Yannick Renaud[2], Sébastien Brunet[2], Genevieve Fortin[1], Jacques Belaiche[4], Alain Bitton[7], Daniel Gaudet[8], Albert Cohen[9], John D. Rioux[10], Ian D.R. Arnott[6], Paul Rutgeerts[11], Jack Satsangi[6], Severine Vermeire[11], Thomas J. Hudson[3] and Denis Franchimont[1,12]

[1] Department of Gastroenterology, McGill University Health Centre and [2] McGill University and Génome Québec Innovation Centre, 740 Dr. Penfield Avenue, Montréal, Québec H3A 1A4, Canada; [3] Ontario Institute for Cancer Research, MaRS Centre, South Tower, 101 College Street, Suite 800, Toronto, Ontario M5G 0A3, Canada; [4] Division of Gastroenterology, CHU, Université de Liège, 13 avenue de l'Hôpital Bât. B35, 4000 Liège, Belgium; [5] Mount Sinai Hospital IBD Centre, University of Toronto, 441-600 University Avenue, Toronto, Ontario M5G 1X5, Canada; [6] Gastrointestinal Unit, Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Edinburgh, United Kingdom; [7] Division of Gastroenterology, McGill University Health Centre, Royal Victoria Hospital, 687 Pine Avenue West, Montréal, Québec H3A 1A1, Canada; [8] Université de Montréal, Centre de médecine génique communautaire de l'Université de Montréal, Hôpital universitaire de Chicoutimi, 555 boulevard de l'Université, Chicoutimi, Québec G7H 2B1, Canada; [9] Jewish General Hospital, 3755 chemin de la Côte-Sainte-Catherine, Montréal, Québec H3T1E2, Canada; [10] Université de Montréal and Montreal Heart Institute, Research Center, 5000 rue Bélanger, Montréal, Québec H1T 1C8, Canada; [11] Division of Gastroenterology, University Hospital Gasthuisberg, 49 Herestraat, 3000 Leuven, Belgium; [12] Department of Gastroenterology, Erasme Hospital, ULB, 808 Lennik road, 1070 Brussels, Belgium.

## 3.1 Rationale

We conceived the following candidate gene study, focusing on the *MEFV* gene, based on the significant association results between the *NLRP3* region and CD susceptibility reported in Chapter 2. *MEFV* encodes for the protein pyrin, and although it is not part of the CATERPILLER family like NOD2 and NALP3, it shares the PYD domain found in NALP proteins. As explained in the introduction, pyrin has been reported to dampen down the NALP3 inflammasome activity (Chae et al. 2006) andto regulate caspase-1 activation and IL-1$\beta$ release, two key mediators of inflammation. Moreover, like it is the case for NOD2 and NALP3, mutations in *MEFV* are responsible for a rare auto-inflammatory syndrome called familial Mediterranean fever. Finally, as explained in the introduction, several clinical and epidemiological evidences linking *MEFV* to IBD further supported this candidate gene selection. We thus evaluated the contribution of *MEFV* to the susceptibility of CD and UC.

# 3.2 Genetic variation in the familial Mediterranean fever gene (*MEFV*) and risk for inflammatory bowel disease

## 3.2.1 Abstract

**Background:** The familial Mediterranean fever (FMF) gene (*MEFV*) encodes pyrin, a major regulator of the inflammasome platform controlling caspase-1 activation and IL-1$\beta$ processing. Pyrin has been shown to interact with the gene product of *NLRP3*, NALP3, also an important active member of the inflammasome. The *NLRP3* region was recently reported to be associated with Crohn's disease (CD) susceptibility. We therefore sought to evaluate *MEFV* as an IBD susceptibility gene. **Material and Results:** *MEFV* colonic mucosal gene expression was significantly increased in experimental colitis mice models (TNBS $p<0.00031$; DSS $p<0.006$), in biopsies from CD ($p<0.020$) and severe Ulcerative colitis (UC) patients ($p<0.008$). Comprehensive genetic screening of the *MEFV* region in the Belgian exploratory sample set (512 CD trios, 113 UC trios, 239 CD cases, 96 UC cases, and 107 healthy controls) identified SNPs located in the *MEFV* 5' haplotype block that were significantly associated with UC (rs224217; $p=0.002545$; A allele frequency: 56% cases, 45% controls), while no associations were observed with CD. Sequencing and subsequent genotyping of variants in exonic and promoter regions located in this associated haplotype block identified three synonymous (D102D, G138G, A165A) and one non-synonymous variants (R202Q) located in *MEFV* exon 2 that were significantly

associated with UC (R202Q: $p$=0.000532; A allele frequency: 32% in cases, 23% in controls). No consistent associations were observed in additional Canadian (226 CD trios, 89 UC trios) and Scottish (521 UC, 300 controls) sample sets. We note that R202Q showed marginal association ($p$=0.012; G allele frequency: 82% in cases, 70% in controls) in the Canadian sample, but with a different risk allele. **Conclusion**: This study suggests that common variants in the *MEFV* region do not contribute to CD and UC susceptibility.

### 3.2.2 Introduction

Crohn's disease (CD) and Ulcerative colitis (UC) are multifactorial and heterogeneous chronically relapsing inflammatory bowel diseases (IBD) that are thought to result from a dysregulated mucosal immune response to gut lumen bacterial antigens in a genetically susceptible host (Podolsky 2002). A recent meta-analysis of three large CD genome-wide association studies (GWAS) has reported that well-established associations with CD only accounts for approximately 20% of the genetic variance observed in CD, suggesting that additional genetic contributions have yet to be discovered (Barrett et al. 2008). Indeed, with the exception of variations within the *NOD2* and *IL23R* genes, established susceptibility alleles have been reported to have relatively modest effects (Barrett et al. 2008). Using a candidate gene approach, *NLRP3* was recently identified as a novel CD susceptibility locus (Villani et al. 2009) that had been missed by previously published GWAS (Duerr et al. 2006; Hampe et al. 2007; Rioux et al. 2007; Libioulle et al. 2007; Parkes et al. 2007; Wellcome Trust

Case Control Consortium 2007a). *NLRP3* (previously known as *CIAS1* or *NALP3*) is part of the CATERPILLER (Ting et al. 2006) gene family and mutations in some of these genes have been shown to result in severe auto-inflammatory diseases (AIDs) like HPFS (Milhavet et al. 2008). AIDs represent a spectrum of diseases characterized by recurrent episodes of seemingly unprovoked inflammation that, unlike autoimmune disorders, lack the production of high-titre auto-antibodies or antigen-specific T cells (Mariathasan et al. 2007). Gain-of-function mutations in *NLRP3* are associated with three hereditary periodic fever syndromes, namely Muckle-Wells syndrome, familial cold autoinflammatory syndrome, and neonatal-onset multisystem inflammatory disease (Agostini et al. 2004; Mariathasan et al. 2007). Importantly, *NLRP3* encodes NALP3 (also known as cryopyrin), a protein playing a key role in controlling the inflammasome, which is a critical molecular platform regulating caspase-1 activation and interleukin (IL)-1$\beta$ processing, two key mediators of inflammation (Mariathasan et al. 2007; Pétrilli et al. 2007a; Martinon et al. 2007).

Recently, the B30.2 (also known as SPRY) domain of pyrin, which is encoded by the *MEFV* gene (Chr.16 p13.3; GenBank NM_000243; OMIM 608107), has been reported to interact and modulate the activity of several inflammasome components, including NALP3, caspase-1 and, its substrate, pro-IL-1$\beta$ (Chae et al. 2006; ; Papin et al. 2007; Chae et al. 2008). Interestingly, *MEFV* missense mutations are implicated in the familial Mediterranean fever (FMF), which is another AID (Pras et al. 1992; The International FMF Consortium 1997; The French FMF Consortium 1997; Bernot et al. 1998).

Additional clinical and epidemiological evidences supported *MEFV* as a potential IBD candidate gene. First, IBD and FMF share common clinical and biologic features. They are both inflammatory disorders characterized by the same chronic relapsing behavior infiltration by neutrophils at the site of injury, and abnormal regulation of apoptosis (Lichtenberger et al. 2004; McDermott 2004). Moreover, FMF affects mainly ethnic groups surrounding the Mediterranean Sea (e.g. non-Ashkenazi Jews, Armenians, Turks, and Arabs) and two small cohort studies from that area have reported a higher prevalence of IBD with particularly severe symptoms in FMF non-Ashkenazi Jewish patients, suggesting possible common underlying mechanisms of inflammation (Cattan et al. 2000; Fidder et al. 2002). Additional studies have suggested that *MEFV* rare missense causative mutations have a potential modifying effect in IBD patients (Fidder et al. 2005; Karban et al. 2005; Giaglis et al. 2006; Sari et al. 2008).

Hence, the clinical and epidemiological evidences linking IBD to the *MEFV* gene, together with the similarities of the *MEFV* and *NLRP3* gene products (pyrin and NALP3, respectively) that share the same signaling pathway, prompted us to hypothesize that *MEFV* could also contribute to CD susceptibility and perhaps to UC as well. We therefore explored whether *MEFV* expression would be regulated in IBD experimental models of colitis and whether *MEFV* variants would be associated with CD and/or UC.

### 3.2.3 Methods

**I. Expression Study**

**Tissue collection.** *Animal*: Male BALB/c mice (between 6 and 8 weeks of age), obtained from Charles River Laboratory (St-Constant, Quebec, Canada), were maintained under conventional housing conditions. Colitis was either induced by rectal instillation of 2.5mg of Trinitrobenzene sulfonate (TNBS) (Sigma Aldrich Canada Ltd., Ontario, Canada) or mice were fed with 5% Dextran Sulfate Sodium (DSS) (MP Biomedicals, OH, USA) using modified protocols previously reported (Okayasu et al. 1990; Neurath et al. 1995). All mice were handled according to institutionally recommended animal care guidelines and all experiments were approved by the Animal Studies Ethics Committee of McGill University. *Human*: Colonic biopsy specimens were obtained from CD patients in ulcerated (severe involvement) and non-ulcerated (mild involvement) mucosa, from UC patients in inflamed areas as defined by Mayo sub-endoscopic criteria, and from patients who underwent colorectal cancer screening (CD n=16; UC n=17; Controls n=25). The study was approved by the Institutional Review Board and informed consent was obtained from all participating subjects.

**RNA extraction and real-time quantitative PCR.** Biopsies preserved in RNALater (Qiagen, Ontario, Canada) were homogenized and total RNA was extracted using TRIZOL (Invitrogen, Ontario, Canada). First strand cDNA was synthesized using the cDNA Archive Kit (Applied Biosystems (ABI), CA, USA) with Multiscribe Reverse Transcriptase and Random primers. For both human and mice, primers and probes were ordered from Assay-on-demand ABI

catalogue (mice *MEFV*: Mm00490260_gl; human *MEFV*: Hs00165145_ml; *18S*: 4319413E). Quantitative real-time PCR was performed using ABI prism 7900 sequence detection system based on the 5' nuclease assay (Holland et al. 1991), and quantified using ABI's comparative Ct method. Statistical analyses were performed using the Wilcoxon Signed Ranks test to evaluate tissue (mice and human) expression differences (SPSS, Version 11.5, USA).

**II. Genetic Study**

**Subjects.** Three main sample sets, coming from 5 different centers, were analyzed in this study (see detailed descriptions, **Table 3.1**), comprising a total of 738 CD trios, 202 UC trios, 521 UC cases, and 300 controls. All protocols were consented by the Institutional Review Board of each institution that sent samples. All patients were recruited through specialized hospitals, academic centres, and practitioners. Inflammatory bowel disease (IBD) specialists involved in this study and in most cases, the CD or UC clinical subtype was assigned using standard criteria (Lennard-Jones JE 1989; Silverberg et al. 2005). Those defined as "indeterminate colitis" were excluded from the study. In all the participating centres, the diagnosis of IBD was made by IBD specialists after fulfilling standard clinical, radiological, endoscopic, and pathology criteria[1] that required (1) one or more of the following symptoms: diarrhea, rectal bleeding, abdominal pain, weight loss, fever or complicated perianal disease; (2) occurrence of symptoms on two or more occasions in the past or ongoing symptoms of at least 4-6 weeks' duration; (3) evidence of inflammation, strictures or fistula from

radiological, endoscopic, and histological evaluation (with some specific CD characteristics); (4) exclusion of all other diagnosis besides CD.

Belgian subjects (referred as combined Belgian cohort), all of European descent, were used for the exploratory phase and these came from Center 1 (University of Liege, Belgium) and Center 2 (University of Leuven, Belgium). The replication cohort consisted of Canadian subjects (referred as combined Canadian cohort), coming from Centers 3 and 4, and Scottish subjects. Center 3 was composed of subjects collected from multiple sites in the province of Quebec (Canada), which were all of European ancestry, and included 38 probands of Ashkenazi Jewish ancestry and 6 of Sephardic Jewish ancestry. Center 4 included subjects collected from multiple sites in Toronto (Canada). The vast majority of probands were of European ancestry, including 40 probands of Ashkenazi Jewish ancestry, and 15 probands were of non-European ancestry. Finally, cases and matched controls from center 5 were collected from multiple sites in Edinburgh (Scotland). The majority of cases were of European ancestry, including 1 case of Jewish ancestry, except for 14 non-European cases and 11 non-European controls. Center 5 will be referred to as the Edinburgh cohort. Overall, only samples of European origin were included in the analysis.

**Genotyping.** SNP genotyping was performed using either a sequencing approach, the SNPstream UltraHigh Throughput Genotyping System (Orchid Biosciences, NJ, USA) (Bell et al. 2002), the Fluorescent Polarization Template-Directed Dye Terminator (FP-TDI) system (PerkinElmer, MA, USA) (Hsu et al. 2001), or the Taqman 5' exonuclease assay (ABI, CA, USA) (rs2242217: assay

C_2394717_20) (Whitcombe et al. 1998). Also, since 2 non-synonymous variants (E148Q, R202Q) as well as three synonymous variants (D102D, G138G, A165A) were observed in *MEFV* exon 2 during sequencing screening, genotyping of these variants was done using a sequencing approach. All primers and probes are available in **Supplementary Table 3.2**.

We used 4 quality checks to determine whether SNPs were reliable and informative: (1) genotyping efficiency (>95%), (2) compliance to Hardy-Weinberg equilibrium ($p>0.01$), (3) Mendelian inheritance, and (4) population minor allele frequency (>5%).

**Sequencing.** Sequencing was performed on an ABI 3730 DNA Sequencer according to standard protocols. Sequence traces were assembled and analyzed using a modified version of the PolyPhred software package. Primer sequences are available in **Supplementary Table 3.3**. Our sequences were compared to the annotated sequences (build 35, hg17) of 25 exons and promoter regions of *MEFV*, *AK096958*, *ZNF263*, *TIGD7* and *ZNF75A* to identify novel variants.

**Statistical analysis.** To evaluate *MEFV* as a candidate gene for IBD, we divided the study into an exploratory and replication phase. Mendelian errors and departures from Hardy-Weinberg equilibrium were assessed using MERLIN v0.9.10. Measures of pairwise LD between SNPs ($D'$ and $r^2$) were computed using Haploview v3.32 (Barrett et al. 2005). Tests of association were performed using the likelihood methods implemented in UNPHASED v3.0.10 (Dudbridge et

al. 2008), which can analyze samples of nuclear families, unrelated subjects or a combination of both.

### 3.2.4 Results

**Increased MEFV expression in IBD experimental models of colitis and in colonic mucosa from IBD patients.**

To further support our candidate gene selection, we first evaluated *MEFV* expression in different models of experimental colitis. *MEFV* expression is significantly increased TNBS ($p<0.000310$) **(Figure 3.1A)** and in DSS ($p<0.006$) **(Figure 3.1B)** colitis mice models as compared to colonic tissues from control mice, considered as the arbitrary baseline = 1. In human tissue biopsies, *MEFV* expression is significantly increased in ulcerated colonic mucosa from CD patients compared to healthy controls ($p<0.020$) **(Figure 3.1C)**, but not from UC patients ($p<0.265$) **(Figure 3.1D)**. However, stratification by severity of disease, as defined by the presence of endoscopic lesions (CD) and Mayo sub-endoscopic criteria (UC), shows that *MEFV* expression is upregulated in both severely affected CD ($p<0.001$) **(Figure 3.1C)** and UC ($p<0.008$) **(Figure 3.1D)** patients compared to healthy controls.

**MEFV mutations in exon 10 do not contribute to CD and UC susceptibility**

FMF is thought to be secondary to missense mutations in *MEFV*. Although these mutations are found throughout *MEFV*, five sequence alterations in *MEFV* represent the majority of FMF chromosomes, four of which are

clustered in exon 10 (i.e. *M680I*, *M694V*, *M694I*, and *V726A*) (Milhavet et al. 2008; Pras et al. 1992; The International FMF Consortium 1997; The French FMF Consortium 1997; Bernot et al. 1998). This exon was sequenced in 47 CD patients, 47 UC patients and 94 controls. Only 3 individuals (2 CD and 1 UC patients) carrying 3 different mutations (*R652H*, *M694V*, and *V726A* respectively) were observed. None of the exon 10 variants could link IBD with *MEFV*. The low frequency of these observed mutations confirms previous reports studying the prevalence and association of these mutations in IBD patients (Cattan et al. 2000; Fidder et al. 2002; Fidder et al. 2005; Karban et al. 2005; Giaglis et al. 2006; Sari et al. 2008).

**Common SNPs located in MEFV 5' region are significantly associated with IBD in the Belgian exploratory sample set.**

We subsequently sought to evaluate *MEFV* common variants for their association with IBD. We first investigated a 98kb region spanning 16p13.3 (3220975-3318978) (NCBI Build 35, hg17), including *MEFV* (14.6kb). SNPs were selected using a pairwise tagging approach (de Bakker et al. 2005), using data from HapMap Public Release #22 ($r^2 \geq 0.8$) (minor allele frequency (MAF) $\geq$ 0.05). The SNP panel was enriched with SNPs located in functional domains, as well as with SNPs selected from dbSNP Build126 in regions with lower coverage. A total of 30 informative SNPs in the *MEFV* region were genotyped in 512 Belgian CD trios and 113 UC trios (**Table 3.1**).

As previously reported, we observe 2 regions of high linkage disequilirium (LD) separated by a recombination "hotspot" within *MEFV* intron 2

(**Figure 3.2**) (Aldea et al. 2004). The 5' haplotype block includes the promoter to *MEFV* intron 2 region, three known genes (*ZNF263*, *TIGD7*, *ZNF75A*), and one hypothetical gene (*AK096958*). The 3' haplotype block spans *MEFV* intron 2 to the 3' UTR region of *MEFV*, as well as the region encompassing the promoter to intron 4 of *ZNF200* (**Figure 3.2**).

The analysis of Belgian CD trios only revealed nominally significant associations with the *major* alleles of three SNPs located in the region covering the *MEFV* promoter to intron 2 and its 5' flanking region (rs182674: $p$=0.02938; rs224217: $p$=0.05240; rs224231; $p$=0.03935; rs6501170: $p$=0.01552) (**Figure 3.2A, Supplementary Table 3.1**).

The analysis of Belgian UC trios revealed significant associations (lowest $p$=0.000689 for rs224217) with the *minor* alleles of eight SNPs (i.e. rs1149483, rs224243, rs224231, rs224230, rs224226, rs224225, rs182674, and rs224217) located in that same region of high LD (**Figure 3.2B**, **Supplementary Table 3.1**). Contrary to the CD associations, the UC associations remained significant after allowing for the total number of tests performed ($p_{corrected}$=0.018 based on 1000 random permutations in 2 diseases). These associations with UC were also significant across the two Belgian centers (rs224217: $p$=0.0380 [Center 1: Leuven]; $p$=0.0008 [Center 2: Liege]) described in **Table 3.1**.

**Sequencing of coding and promoter regions of genes in the associated 5' haplotype block**

Following the identification of the association signals with CD and UC in the 5' haplotype block, we sequenced the exons and promoter regions (2kb upstream) of all genes located within this region to exclude the involvement of other genes (i.e. *AK096958, ZNF263, TIGD7, ZNF75A*) beside *MEFV*. A total of 10 UC patients, 10 CD patients and 10 healthy controls were sequenced. A total of 9 synonymous and 9 non-synonymous variants were observed. A detailed description of all observed coding variants can be found in **Table 3.2**.

**Augmenting the sample size: MEFV common SNPs do not contribute to CD and UC susceptibility**

***The Belgian samples***

To further validate the Belgian association results, we first augmented the sample size by genotyping additional unrelated Belgian cases and controls (239 CD, 96 UC and 107 shared controls) and combining them with the 512 CD and 113 UC trios. The cases and controls were genotyped for 12 of the 30 exploratory set of SNPs, which were chosen using a pairwise tagging approach (de Bakker et al. 2005) to remove the redundancy among the genotyped markers, given the high level of LD between them. Additionally, all Belgian case-control and trios samples were genotyped for the non-synonymous variants observed in the sequencing experiment, which are described in **Table 3.2**.

When increasing the sample size in the combined Belgian CD case-control-trios analyses, no more significant associations were observed between

the 12 tagging SNPs and CD (**Figure 3.2A**, **Table 3.4**, **Supplementary Table 3.1**). Moreover, no significant associations were observed between all observed non-synonymous coding SNPs described in **Table 3.2** and the combined CD Belgian sample set (**Table 3.4**).

In the combined Belgian UC case-control-trios analyses, the association significance levels observed with the 12 tagging SNPs were similar to the ones observed in the analysis focussing solely on trios (**Figure 3.2B**, **Table 3.3**, **Supplementary Table 3.1**). The UC association analysis of the non-synonymous variants from **Table 3.2** revealed no significant associations with *C310S* (rs220379), previously genotyped, as well as with the rare coding variants *R26Q* (3 carriers), *V534I* (11 carriers), and *N248H* (8 carriers). Similarly, no significant associations with coding variants located in 3' end region of *MEFV* were detected. Yet, significant associations were observed with 4 coding variants (*D102D*, *G138G*, *A165A,* and *R202Q*) located in *MEFV* exon 2. The *minor* alleles (MAF=45%) of the three synonymous variants (*D102D*, *G138G*, *A165A*) were associated with UC in the Belgian combined sample set ($p$=0.00146) (**Table 3.3**); these three variants are in perfect LD with each other and in high LD with tagging SNP rs224217 ($r^2$=0.85). Additionally, the *minor* allele of the non-synonymous SNP *R202Q* was also significantly associated with the UC (R202Q/rs224222: $p$=0.000532; A allele frequency: 32% cases, 23% controls) in the Belgian samples.

To assess the reproducibility of these results, we subsequently evaluated the association of the same 12 tagging SNPs and the 4 coding SNPs located in *MEFV* exon 2 region (i.e. D102D, G138G, A165A and R202Q) in additional

samples from Canada (226 CD trios, 89 UC trios) and Scotland (521 UC, 300 controls) (**Table 3.1**).

***The Canadian samples***

In the Canadian CD sample set, only the minor allele of tagging SNP rs224230, located in the 5' flanking region of *MEFV*, was significantly associated with CD ($p$=0.04387; A allele frequency: 40% cases, 33% controls) (**Table 3.4**). None of the 16 SNPs were consistently replicated in the Belgian and Canadian CD sample sets, and combined analysis of the Belgian-Canadian CD samples (**Table 3.4**) did not bring further insight on additional associations in the region, supporting the exclusion of the *MEFV* region as a risk factor contributing to CD susceptibility.

In the Canadian UC sample set, individual tagging SNPs analysis uncovered one nominally significant associations in the *MEFV* 5' region (rs220379/*C310S* $p$=0.0456; G allele frequency: 81% cases, 70% controls ) and one association with exon 2 coding SNP R202Q (R202Q/rs224222: $p$=0.0117; G allele frequency: 82% cases, 70% controls) (**Table 3.3**). Contrary to the association in the Belgian samples where R202Q *minor* allele was associated with UC susceptibility, the *major* allele of *R202Q* was associated with UC in the Canadian samples, and no associations were observed with the three synonymous variants of exon 2. None of these results replicated the initial significant findings observed in the Belgian exploratory UC sample set (**Figure 3.2**).

***The Scottish samples***

In the third UC case-control cohort from Scotland (Edinburgh) (**Table 3.1**), only the major allele of tagging SNP rs224215 located in *MEFV* intron 2 was significantly associated with UC ($p$= 0.0406; A allele frequency: 63% cases, 58% controls). None of the associations observed in the Belgian and Canadian UC analyses were replicated in the Scottish sample set (**Table 3.3**). No SNPs were significantly associated in the combined analysis of all Belgian-Canadian-Scottish UC samples (**Table 3.3**).

## 3.2.5 Figures

**Figure 3.1: Level of MEFV mRNA expression in inflamed colitis tissues compared to healthy controls**

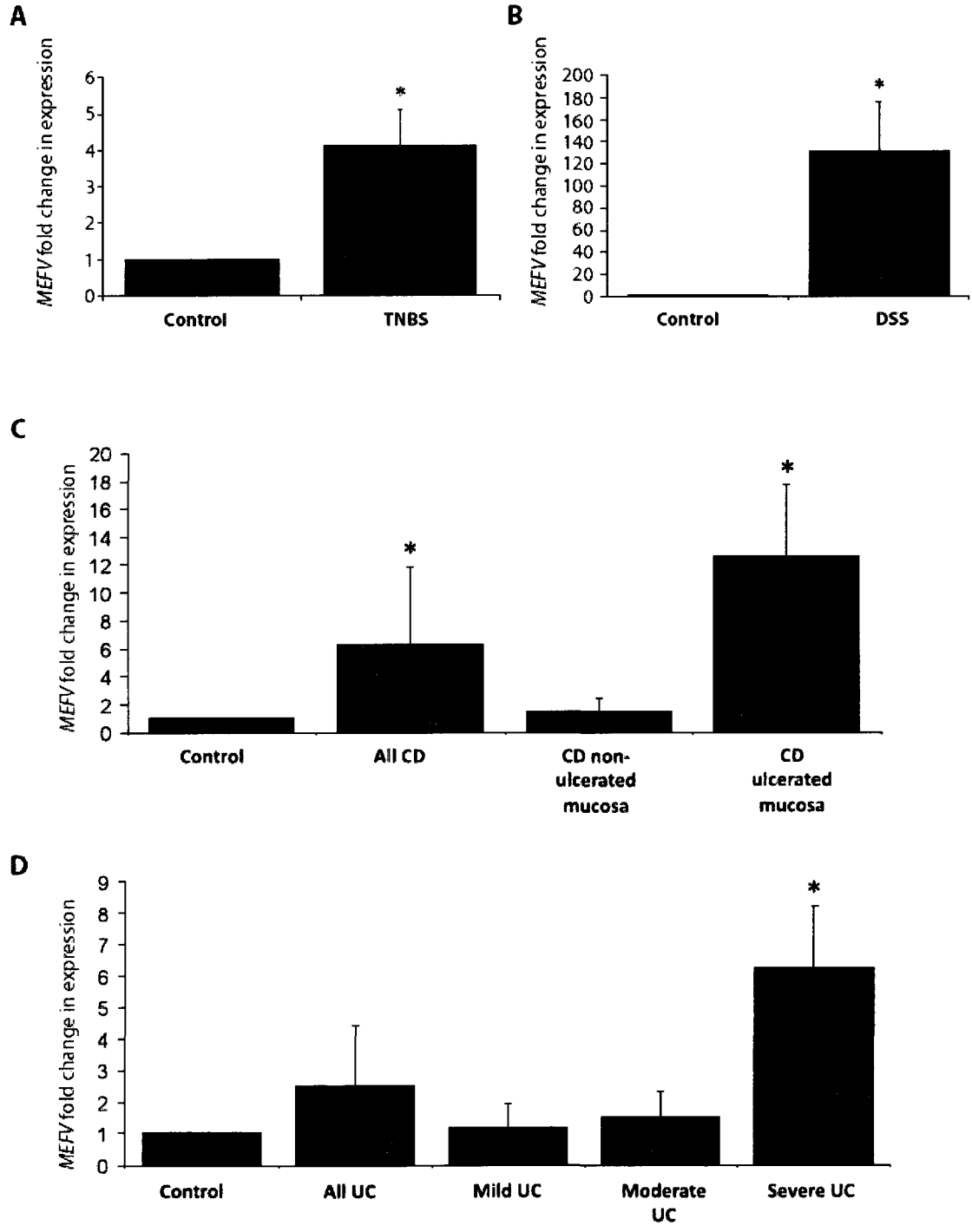


**Figure 3.1:** (a, b) *MEFV* expression was assessed by quantitative real-time PCR in (a, b) healthy mice colons (n=7), in colons from (a) acute TNBS (n=8) and (b) DSS (n=4) colitis mice models, as well as in (c, d) healthy human (n=25), (c) CD (n=16) and (d) UC (n=17) colonic specimens. *MEFV* expression appeared to correlate with disease severity in both CD (c: 9 CD with mild inflammation and 7 CD with severe inflammation) and UC (d: 6 UC with mild inflammation, 7 UC with moderate inflammation and 4UC with severe inflammation) samples. (a-d) Expression was normalized to *18S* gene expression and each bar represents the mean value ± S.D. (*) = $p < 0.05$ compared to healthy colonic specimens, considered as the arbitrary baseline = 1.

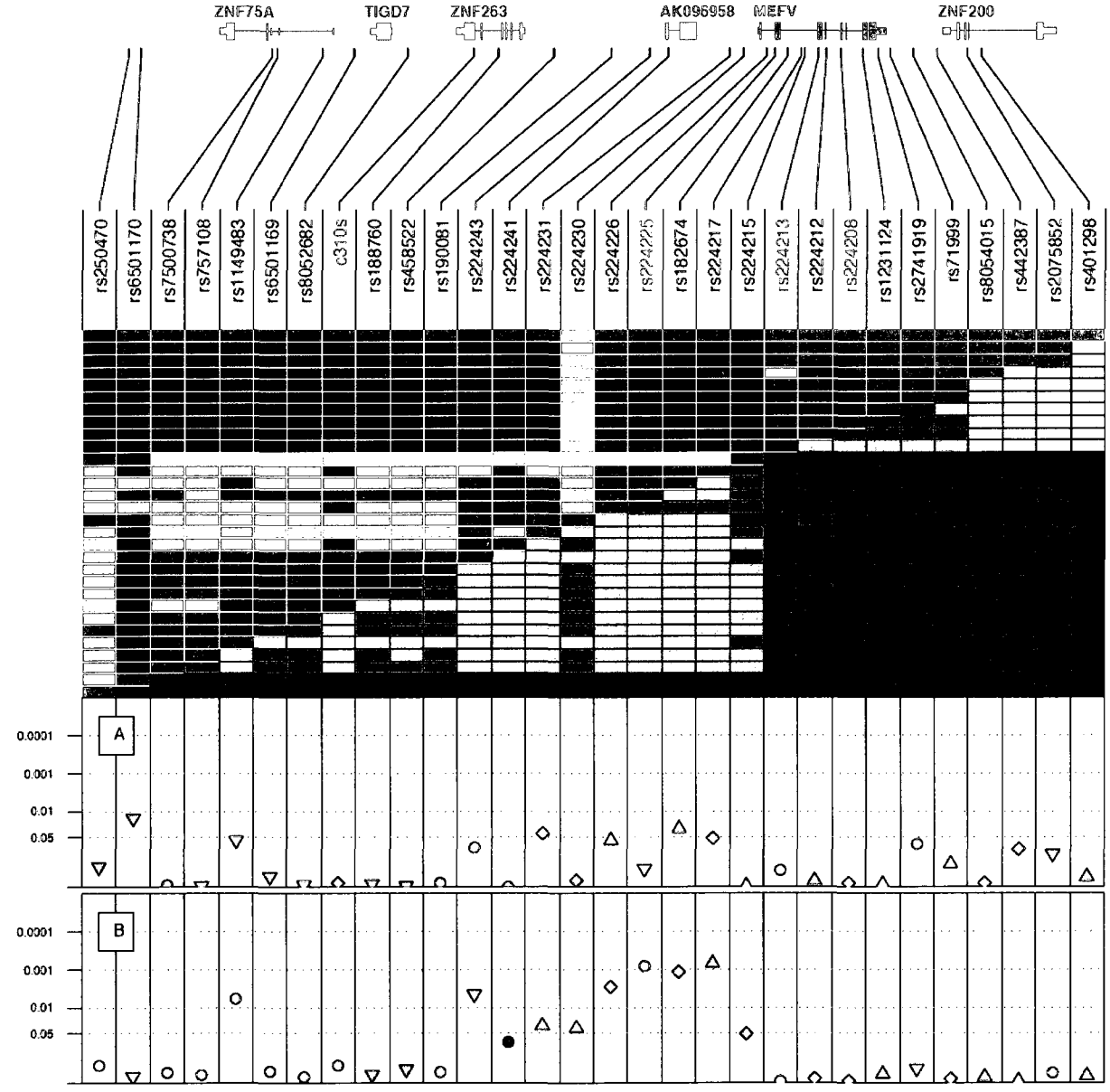



**Figure 3.2: Association results of the SNP panel screened in Belgian IBD exploratory cohorts.** Shown above are the SNPs with their positions in the genes and the LD structure between them. SNPs in red are exonic. The upper left portion is *D'* and the lower right portion is $r^2$. In the lower panel are reported the results from association analysis of the Belgian CD (a), and the Belgian UC (b) samples. Level of significance can be found on the scale located at the bottom left of the figure. *P* value of individual alleles are reported, where the symbols represent the associated allele (Δ=T, □=C, ○=A, ◊=G) and the color scheme represents the allele frequency.

### 3.2.6 Tables

**Table 3.1: Subjects examined as part of the *MEFV* candidate gene study**

| | Numbers of affected subjects | | | |
|---|---|---|---|---|
| | **IBD** | **CD** | **UC** | **Control** |
| **Exploratory Combined Belgian Cohort** | | | | |
| 1- Leuven Trios | 435 | 356 | 79 | N/A |
| 2a- Liege Trios | 190 | 156 | 34 | N/A |
| 2b- Liege C/C[a] | 335 | 239 | 96 | 107 |
| Total Exploratory Cohort | 960 | 751 | 209 | 107 |
| **Replication Cohorts :** | | | | |
| **Canadian Combined Cohort** | | | | |
| 3- Quebec Trios[b] | 125 | 95 | 30 | N/A |
| 4- Toronto Trios[c] | 190 | 131 | 59 | N/A |
| Total Canadian Cohort | 315 | 226 | 89 | N/A |
| **Scottish Cohort** | | | | |
| 5- Edinburth (Scottish) C/C[a,d] | 521 | 0 | 521 | 300 |
| Total Replication Cohort | 836 | 226 | 610 | 300 |

[a] C/C refers to case control sample set.

[b] Includes 27 CD trios and 17 UC trios of Jewish ancestry

[c] Includes 26 CD trios and 14 UC trios of Jewish ancestry, as well as 15 non-European probands (these were excluded from the analysis)

[d] Includes 1 UC case of Jewish ancestry; includes also 14 cases and 11 controls of non-European descent, which were excluded from the analysis.

**Table 3.2: Coding variants unovered in *MEFV* 5' block**

| Base Change[a] | Amino Acid Change | dbSNP#[b] | Gene | Location |
|---|---|---|---|---|
| c.77G>A | R26Q | Novel | *TIGD7* | exon1 |
| c.1600G>A | V534I | Novel | *ZNF263* | exon 6 |
| c.929G>C | C310S | rs220379 | *ZNF263* | exon 6 |
| c.742A>C | N248H | Novel | *AK096958* | exon1 |
| c.306T>C | D102D | rs224225 | *MEFV* | exon 2 |
| c.414A>G | G138G | rs224224 | *MEFV* | exon 2 |
| c.442G>C | E148Q | rs3743930 | *MEFV* | exon 2 |
| c.495C>A | A165A | rs224223 | *MEFV* | exon 2 |
| c.605G>A | R202Q | rs224222 | *MEFV* | exon 2 |
| c.942C>T | R314R | rs224213 | *MEFV* | exon 3 |
| c.1105C>T | P369S | rs11466023 | *MEFV* | exon 3 |
| c.1223G>A | R408Q | rs11466024 | *MEFV* | exon 3 |
| c.1422G>A | E474E | rs224208 | *MEFV* | exon 5 |
| c.1428A>G | Q476Q | rs224207 | *MEFV* | exon 5 |
| c.1530T>C | D510D | rs224206 | *MEFV* | exon 5 |
| c.1648C>G | P550A | Novel | *MEFV* | exon 7 |
| c.1764G>A | P588P | rs1231122 | *MEFV* | exon 9 |
| c.2118G>A | P706P | rs2234939 | *MEFV* | exon 10 |

[a] According to the cDNA coding sequence, with +1 from the A of the initiating ATG. Reference sequences are NM_033208 (*TIGD7*), NM_005741 (*ZNF263*), cDNA *AK096958* and NM_000243 (*MEFV*); [b] dbSNP: www.ncbi.nlm.nih.gov/SNP

## Table 3.3: Association results of tagging SNPs and *MEFV* exon 2 coding SNPs in the UC sample sets

| | | | Belgian UC | | | | Canadian UC | | | | Edinburgh UC | | | | Combined UC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | SNP | Position | Allele | Frequency Cases | Frequency Controls | *P* value | Allele | Frequency Cases | Frequency Controls | *P* value | Allele | Frequency Cases | Frequency Controls | *P* value | Allele | Frequency Cases | Frequency Controls | *P* value |
| 1 | rs250470 | 3318978 | 2 | 0.75 | 0.73 | 0.8473 | 4 | 0.26 | 0.24 | 0.6681 | 4 | 0.21 | 0.21 | 0.8969 | 2 | 0.77 | 0.77 | 0.4075 |
| 2 | rs6501170 | 3317583 | 2 | 0.17 | 0.16 | 0.7423 | 4 | 0.87 | 0.82 | 0.3022 | 2 | 0.16 | 0.16 | 0.9747 | 2 | 0.16 | 0.16 | 0.7213 |
| 3 | C310S | 3279436 | 2 | 0.26 | 0.23 | 0.1333 | 3 | 0.81 | 0.70 | 0.0457 | 2 | 0.25 | 0.24 | 0.8262 | 3 | 0.75 | 0.75 | 0.7041 |
| 4 | rs190081 | 3263570 | 2 | 0.27 | 0.24 | 0.1101 | 4 | 0.82 | 0.74 | 0.0881 | 2 | 0.24 | 0.24 | 0.8593 | 2 | 0.24 | 0.24 | 0.6248 |
| 5 | rs224243 | 3259194 | 4 | 0.53 | 0.43 | 0.005196 | 2 | 0.57 | 0.54 | 0.7098 | 4 | 0.48 | 0.48 | 0.9977 | 4 | 0.49 | 0.46 | 0.0768 |
| 6 | rs224241 | 3257037 | 2 | 0.26 | 0.23 | 0.0490 | 4 | 0.81 | 0.71 | 0.0727 | 4 | 0.77 | 0.77 | 0.9255 | 4 | 0.77 | 0.76 | 0.7605 |
| 7 | rs224230 | 3248358 | 1 | 0.71 | 0.66 | 0.1979 | 3 | 0.36 | 0.33 | 0.5148 | 3 | 0.36 | 0.35 | 0.7233 | 1 | 0.66 | 0.66 | 0.8716 |
| 8 | rs224225 | 3244763 | 2 | 0.56 | 0.45 | 0.001463 | 4 | 0.59 | 0.54 | 0.3385 | 2 | 0.46 | 0.45 | 0.7062 | 2 | 0.48 | 0.45 | 0.1897 |
| 9 | rs224224 | 3244655 | 3 | 0.55 | 0.45 | 0.002625 | 1 | 0.59 | 0.54 | 0.3385 | 3 | 0.46 | 0.45 | 0.7062 | 3 | 0.48 | 0.45 | 0.2079 |
| 10 | rs224223 | 3244574 | 1 | 0.55 | 0.45 | 0.001981 | 2 | 0.59 | 0.54 | 0.3385 | 1 | 0.46 | 0.44 | 0.6245 | 1 | 0.48 | 0.45 | 0.1686 |
| 11 | R202Q | 3244464 | 1 | 0.32 | 0.23 | 0.000532 | 3 | 0.82 | 0.70 | 0.0117 | 3 | 0.74 | 0.74 | 0.9823 | 1 | 0.27 | 0.26 | 0.1789 |
| 12 | rs224217 | 3241758 | 1 | 0.56 | 0.45 | 0.002545 | 3 | 0.57 | 0.55 | 0.7179 | 3 | 0.51 | 0.50 | 0.8781 | 1 | 0.51 | 0.48 | 0.0694 |
| 13 | rs224215 | 3241361 | 3 | 0.38 | 0.37 | 0.4363 | 1 | 0.68 | 0.67 | 0.7855 | 1 | 0.63 | 0.58 | 0.0406 | 1 | 0.63 | 0.61 | 0.6790 |
| 14 | rs1231124 | 3234679 | 1 | 0.50 | 0.50 | 0.5949 | 1 | 0.49 | 0.44 | 0.3824 | 1 | 0.45 | 0.45 | 0.9495 | 1 | 0.47 | 0.46 | 0.9719 |
| 15 | rs442387 | 3226119 | 3 | 0.59 | 0.53 | 0.0595 | 3 | 0.56 | 0.54 | 0.7255 | 3 | 0.50 | 0.50 | 0.9109 | 3 | 0.53 | 0.52 | 0.8027 |
| 16 | rs401298 | 3220975 | 1 | 0.46 | 0.42 | 0.2741 | 1 | 0.43 | 0.36 | 0.3107 | 3 | 0.61 | 0.59 | 0.3107 | 3 | 0.59 | 0.59 | 0.8714 |

**Table 3.4: Association results of tagging SNPs and *MEFV* exon 2 coding SNPs in CD sample sets**

| | | | Belgian CD | | | | Canadian CD | | | | Combined CD | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | SNP | Position | Allele | Frequency Cases | Frequency Controls | *P* value | Allele | Frequency Cases | Frequency Controls | *P* value | Allele | Frequency Cases | Frequency Controls | *P* value |
| 1 | rs250470 | 3318978 | 4 | 0.27 | 0.26 | 0.5166 | 2 | 0.72 | 0.72 | 1.0000 | 4 | 0.27 | 0.26 | 0.6874 |
| 2 | rs6501170 | 3317583 | 4 | 0.86 | 0.83 | 0.1133 | 4 | 0.87 | 0.82 | 0.0533 | 4 | 0.86 | 0.83 | 0.0179 |
| 3 | C310S | 3279436 | 3 | 0.77 | 0.76 | 0.6680 | 3 | 0.76 | 0.71 | 0.1839 | 3 | 0.76 | 0.74 | 0.2080 |
| 4 | rs190081 | 3263570 | 4 | 0.76 | 0.76 | 0.9732 | 4 | 0.75 | 0.71 | 0.1718 | 4 | 0.76 | 0.74 | 0.3491 |
| 5 | rs224243 | 3259194 | 2 | 0.55 | 0.53 | 0.5570 | 4 | 0.42 | 0.42 | 0.8846 | 2 | 0.56 | 0.55 | 0.8983 |
| 6 | rs224241 | 3257037 | 4 | 0.77 | 0.75 | 0.2282 | 4 | 0.76 | 0.73 | 0.3169 | 4 | 0.77 | 0.74 | 0.0964 |
| 7 | rs224230 | 3248358 | 1 | 0.66 | 0.66 | 0.7677 | 3 | 0.40 | 0.33 | 0.0439 | 3 | 0.36 | 0.34 | 0.6047 |
| 8 | rs224225 | 3244763 | 4 | 0.52 | 0.51 | 0.8556 | 4 | 0.57 | 0.55 | 0.6298 | 4 | 0.53 | 0.52 | 0.7230 |
| 9 | rs224224 | 3244655 | 1 | 0.51 | 0.51 | 0.7661 | 1 | 0.57 | 0.55 | 0.6298 | 1 | 0.53 | 0.53 | 0.6503 |
| 10 | rs224223 | 3244574 | 2 | 0.52 | 0.51 | 0.7818 | 2 | 0.57 | 0.55 | 0.6298 | 2 | 0.53 | 0.53 | 0.6663 |
| 11 | R202Q | 3244464 | 1 | 0.27 | 0.26 | 0.4749 | 3 | 0.75 | 0.70 | 0.0885 | 3 | 0.74 | 0.72 | 0.7535 |
| 12 | rs224217 | 3241758 | 3 | 0.53 | 0.51 | 0.8638 | 3 | 0.57 | 0.56 | 0.8290 | 3 | 0.54 | 0.53 | 0.8713 |
| 13 | rs224215 | 3241361 | 3 | 0.39 | 0.39 | 0.7628 | 3 | 0.38 | 0.37 | 0.6582 | 3 | 0.39 | 0.38 | 0.9612 |
| 14 | rs1231124 | 3234679 | 1 | 0.48 | 0.44 | 0.0822 | 1 | 0.49 | 0.42 | 0.0563 | 1 | 0.48 | 0.43 | 0.0134 |
| 15 | rs442387 | 3226119 | 3 | 0.54 | 0.52 | 0.2781 | 3 | 0.56 | 0.51 | 0.1526 | 3 | 0.55 | 0.51 | 0.1059 |
| 16 | rs401298 | 3220975 | 1 | 0.40 | 0.40 | 0.8400 | 1 | 0.41 | 0.37 | 0.2182 | 1 | 0.40 | 0.39 | 0.4089 |

### 3.2.7 Discussion

Genes, such as *MEFV*, in which mutations lead to severe systemic inflammatory diseases, like the AIDs, herald their critical role in the control of inflammation and may represent potential candidates for the onset chronic inflammatory disorders like CD and UC. To further support our candidate gene selection, we integrated results from mice and human colonic expression studies of *MEFV*. We found *MEFV* expression to be significantly increased in both TNBS colitis mice model, which mimics CD-like intestinal inflammation, and in DSS colitis mice model, which mimics UC-like intestinal inflammation. We also found an increase in gene expression in inflamed colonic tissues from both CD and UC patients, correlating with severity of inflammation. Fine mapping of the *MEFV* region revealed two haplotype-blocks with a previously reported recombination hotspot located in *MEFV* intron 2 (Aldea et al. 2004). In our exploratory phase, we report associations between the *minor* allele of variants in the *MEFV* 5' region and UC, as well as associations between the *major* allele of variants located in the same region and CD in the Belgian sample set. Yet, we failed to subsequently validate these associations in the Canadian and Scottish IBD sample sets.

Unlike previous studies (Cattan et al. 2000; Fidder et al. 2002; Fidder et al. 2005; Karban et al. 2005) that have excluded the involvement of *MEFV* in CD pathogenesis by looking at specific rare FMF missense causative mutations clustered in exon 2 and 10 in a relatively small number of CD cases, we carried out a thorough genetic screening of the *MEFV* region in two large CD sample

sets where we observed no association between common variants in the *MEFV* region and CD. Since no variants were consistently associated with CD in the *MEFV* region, we concluded in agreement with previous studies looking at rare FMF causative mutations (Cattan et al. 2000; Fidder et al. 2002; Fidder et al. 2005; Karban et al. 2005) that common variants in the *MEFV* region do not contribute to CD susceptibility.

Upon the observation of significant associations between SNPs located in the 5' block region of *MEFV* and UC in the exploratory phase, we excluded possible coding risk variants and non-*MEFV* genes located in the 5' block by sequencing all the exonic and promoter regions of the five genes in this region and genotyping the uncovered non-synonymous variants. Only coding SNPs within *MEFV* exon 2 (i.e. *D102D*, *G138G*, *A165A*, and *R202Q*) were significantly associated with UC in the Belgian samples. Yet, in the validation phase, none of the synonymous variants were associated with UC in the Canadian samples and we observed opposite *R202Q* (rs224222) allele association with UC in the Belgian and Canadian cohorts, suggesting that *R202Q* was unlikely to be a causative variant contributing to UC susceptibility.

Of interest, several other groups have reported the contribution of *MEFV* and SNPs in the same 5' region to the susceptibility of UC and other chronic inflammatory disorder. For example, *MEFV* exon 2 haplotype CGAA/G (*D102D*, *G138G*, *A165A*, *R202Q*), associated with UC in our study, has also been associated with FMF causative rare mutations in Mediterranean populations (Aldea et al. 2004). A study conducted in a Greek population, focusing precisely on *MEFV* exon 2 and 10, has reported an association between UC and *MEFV*

exon 2 haplotype (Giaglis et al. 2006). Yet, the FMF causative mutations were reported to be associated with another exon 2 hapltoype (i.e. *D102D*, *G138G*, *A165A*, *R202Q*: TACG) in this population, a difference in association that could reflect a distinct founder effect or be consequent of the relative small sample size evaluated in their study (Giaglis et al. 2006). In agreement with our UC preliminary association results, another study reported the association of six SNPs in the *MEFV* region with juvenile idiopathic arthritis (Day et al. 2008). Amongst these SNPs, three of them (i.e. rs224217, rs224225, rs224223) located in the *MEFV* 5' region were also associated in our study. Although of modest size, these studies pointed towards the same *MEFV* region as being involved in inflammatory processes.

Despite the above mentioned studies, our UC exploratory association results need to be interpreted with cautious. Indeed, the associations were observed in two relatively small UC sample sets (totalizing 202 Belgian-Canadian trios, 96 UC cases and 107 healthy controls), and no associations were replicated in our largest UC sample set from Edinburgh (521 UC cases and 300 controls). Additionally, in the combined analyses of all sample sets, we observed no SNPs significantly associated with UC. Given the lack of consistent replication across all our screened sample sets and the lack of SNP associations in the largest sample set (i.e. combined analysis), we conclude that *MEFV* is unlikely to contribute to UC susceptibility.

The fact that mutations in both *MEFV* and *NLRP3* cause auto-inflammatory diseases, that these two genes encode for pyrin domain containing-proteins (Ting et al. 2006) that participate in the same shared signaling pathway

(i.e. the inflammasome) (Chae et al. 2006; Papin et al. 2007; Chae et al. 2008), as well as the recent description of associated common variants in the *NLRP3* region with CD (Villani et al. 2009), prompted us to hypothesize that *MEFV* could equally well be implicated in CD and UC susceptibility. Yet, our results suggests otherwise as we observed no consistent associations with common variants in the *MEFV* region across all our screened sample sets. We thus conclude that *MEFV* does not contribute to CD and UC susceptibility.

### 3.2.8 Accession numbers

GenBank: *NLRP3*, AF054176; *MEFV*, AF018080; *ZNF263*, AC004232; *TIGD7*, AF251050; *ZNF75A,* NM_153028; *ZNF200,* AF060866; *AK096958,* NC_000016

### 3.2.9 Acknowledgement

We acknowledge the IBD patients and their families for their collaboration. The authors would also like to thank A. Montpetit and T. Kwan for their help in the preparation of this manuscript as well as J. Stempak, P. Lepage and A. Verner for their technical support. DF is supported by an infrastructure grant from the Canada Foundation for Innovation (CFI), the Canada Research Chair program, the McGill Research Institute, and the Crohn's & Colitis foundation of Canada (CCFC). ACV is supported by a PhD training award from the Fonds de la Recherche en Santé du Québec (FRSQ). None of the funding agencies had any role in the design and conduct of the study, in the collection, analysis and interpretation of the data or in the preparation, review or approval of the manuscript.

## 3.2.10 Supplementary tables

**Supplementary Table 3.1: Association results of the SNPs genotyped in the exploratory phase in the Belgian CD and UC sample sets**

| | | | Belgian CD | | | | Belgian UC | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| # | SNP ID | Position | Associated Allele | Frequency Cases | Frequency Controls | *P* value | Associated Allele | Frequency Cases | Frequency Controls | *P* value |
| 1 | rs250470 | 3318978 | 4 | 0.27 | 0.25 | 0.2995 | 2 | 0.79 | 0.74 | 0.3446 |
| 2 | rs6501170 | 3317583 | 4 | 0.87 | 0.82 | 0.0155 | 4 | 0.81 | 0.79 | 0.6681 |
| 3 | rs7500738 | 3302610 | 2 | 0.24 | 0.24 | 0.8989 | 2 | 0.27 | 0.24 | 0.5285 |
| 4 | rs757108 | 3302018 | 4 | 0.76 | 0.76 | 0.9491 | 2 | 0.27 | 0.24 | 0.6113 |
| 5 | rs1149483 | 3296835 | 4 | 0.58 | 0.53 | 0.0594 | 2 | 0.55 | 0.40 | 0.0059 |
| 6 | rs6501169 | 3293206 | 4 | 0.77 | 0.76 | 0.5335 | 2 | 0.26 | 0.23 | 0.4919 |
| 7 | rs8052682 | 3287034 | 4 | 0.76 | 0.76 | 0.8951 | 2 | 0.26 | 0.24 | 0.6961 |
| 8 | C310S | 3279436 | 3 | 0.76 | 0.76 | 0.7928 | 2 | 0.27 | 0.22 | 0.3446 |
| 9 | rs188760 | 3276562 | 4 | 0.24 | 0.24 | 0.8468 | 4 | 0.26 | 0.24 | 0.6054 |
| 10 | rs458522 | 3270180 | 4 | 0.24 | 0.24 | 0.9491 | 4 | 0.27 | 0.23 | 0.4382 |
| 11 | rs190081 | 3263570 | 2 | 0.24 | 0.23 | 0.7946 | 2 | 0.27 | 0.24 | 0.5148 |
| 12 | rs224243 | 3259194 | 2 | 0.57 | 0.52 | 0.0920 | 4 | 0.56 | 0.40 | 0.004553 |
| 13 | rs224241 | 3257037 | 2 | 0.22 | 0.22 | 1.0000 | 2 | 0.29 | 0.18 | 0.0840 |
| 14 | rs224231 | 3249980 | 3 | 0.60 | 0.54 | 0.0394 | 1 | 0.57 | 0.42 | 0.0308 |
| 15 | rs224230 | 3248358 | 3 | 0.35 | 0.34 | 0.6872 | 1 | 0.72 | 0.59 | 0.0356 |
| 16 | rs224226 | 3245733 | 1 | 0.57 | 0.52 | 0.0592 | 3 | 0.56 | 0.38 | 0.002867 |
| 17 | rs224225 | 3244763 | 4 | 0.53 | 0.51 | 0.3365 | 2 | 0.58 | 0.40 | 0.00082 |
| 18 | rs182674 | 3243311 | 1 | 0.56 | 0.50 | 0.0294 | 3 | 0.57 | 0.36 | 0.001159 |
| 19 | rs224217 | 3241758 | 3 | 0.56 | 0.51 | 0.0524 | 1 | 0.58 | 0.39 | 0.000689 |
| 20 | rs224215 | 3241361 | 1 | 0.64 | 0.64 | 0.9512 | 3 | 0.42 | 0.29 | 0.0508 |
| 21 | rs224213 | 3239750 | 2 | 0.48 | 0.45 | 0.3673 | 2 | 0.50 | 0.49 | 0.8981 |
| 22 | rs224212 | 3238866 | 1 | 0.47 | 0.45 | 0.6949 | 3 | 0.50 | 0.48 | 0.7728 |
| 23 | rs224208 | 3237182 | 3 | 0.47 | 0.46 | 0.8033 | 3 | 0.52 | 0.51 | 0.8886 |
| 24 | rs1231124 | 3234679 | 1 | 0.47 | 0.46 | 0.8510 | 1 | 0.52 | 0.48 | 0.5860 |
| 25 | rs2741919 | 3232897 | 2 | 0.49 | 0.43 | 0.0752 | 4 | 0.51 | 0.45 | 0.4227 |
| 26 | rs71999 | 3231468 | 1 | 0.48 | 0.45 | 0.2509 | 3 | 0.50 | 0.48 | 0.7773 |
| 27 | rs8054015 | 3228830 | 3 | 0.58 | 0.57 | 0.7987 | 1 | 0.45 | 0.42 | 0.7054 |
| 28 | rs442387 | 3226119 | 3 | 0.57 | 0.52 | 0.1011 | 1 | 0.43 | 0.42 | 0.8886 |
| 29 | rs2075852 | 3222606 | 4 | 0.56 | 0.51 | 0.1379 | 2 | 0.44 | 0.40 | 0.5462 |
| 30 | rs401298 | 3220975 | 1 | 0.43 | 0.41 | 0.5520 | 1 | 0.48 | 0.45 | 0.6616 |

**Supplementary Table 3.2: List of oligos and probes used to perform the genotyping experiments**

| Oligo ID | Primers | Probes | Genotyping Technology | Comments |
|---|---|---|---|---|
| P369S_F | YGGTGCCAGGACTCCCAT | | SNPstream panel I | Below 5% minor allele frequency |
| P369S_R | TGGACCTGCTTCAGGTGG | | SNPstream panel I | Below 5% minor allele frequency |
| P369S_CT_U1 | - | ACGCACGTCCACGGTGATTTAAGGAAGAGCCCGGGAAGCCTAAGC | SNPstream panel I | Below 5% minor allele frequency |
| rs7187449_F | TCCTGGAACATGGAATGG | - | SNPstream panel I | Monomorphic marker |
| rs7187449_R | AAGGCAGCATGTAGGGGT | - | SNPstream panel I | Monomorphic marker |
| rs7187449_CT_U2 | - | GGATGGCGTTCCGTCCTATTTCTTTATACAGACAACTAAGACTGG | SNPstream panel I | Monomorphic marker |
| rs224225_D102D_F | TTCATCATTTTGCATCTGGTT | - | SNPstream panel I | - |
| rs224225_D102D_R | TCTTCAGGCTCCTGGGCT | - | SNPstream panel I | - |
| rs224225_D102D_CT_U2 | - | GGATGGCGTTCCGTCCTATTATTCCACACAAGAAAACGGCACAGA | SNPstream panel I | - |
| rs224241_F | AAATACCACCTGATGTTGGCT | - | SNPstream panel I | - |
| rs224241_R | TGAAATTTTAAAAATGCAGTTGAG | - | SNPstream panel I | - |
| rs224241_CT_U3 | - | CGTGCCGCTCGTGATAGAATAAGAACCAAAGCTTTTCTTGGCAAA | SNPstream panel I | - |
| R42W_F | TTCAAGTTCAAGCTGCAGAAC | - | SNPstream panel I | Below 5% minor allele frequency |
| R42W_R | AGAGTGGCCATCTTCACC | - | SNPstream panel I | Below 5% minor allele frequency |
| R42W_CT_U4 | - | AGCGATCTGCGAGACCGTATGCAGAAGGAGCACTCCAGGATCCCC | SNPstream panel I | Below 5% minor allele frequency |
| V726A_F | TAATAAAGGAGCCTCCCAAGC | - | SNPstream panel I | Below 5% minor allele frequency |
| V726A_R | CCAGAGAAAGAGCAGCTGG | - | SNPstream panel I | Below 5% minor allele frequency |
| V726A_CT_U5 | - | GCGGTAGGTTCCCGACATATGTGGGCATCTTCGTGGACTACAGAG | SNPstream panel I | Below 5% minor allele frequency |
| rs2075852_F | AAATACATTCAAATCCTCAAAGACC | - | SNPstream panel I | - |
| rs2075852_R | ATTTATAGGTCCCTCTTGCTATCC | - | SNPstream panel I | - |

| | | | | |
|---|---|---|---|---|
| rs2075852_CT_U6 | - | GGCTATGATTCGCAATGCTTAACTGAGGTTCACTGACAGCCCTAC | SNPstream panel I | - |
| rs8052921_F | GTGCTGTGTGATCCCATCTC | - | SNPstream panel I | Marker did not work |
| rs8052921_R | TCTGTTGCAAAACGTGTCAC | - | SNPstream panel I | Marker did not work |
| rs8052921_CT_U7 | - | AGGGTCTCTACGCTGACGATTGGTCAGAAAACTCCTCGGAGCTGA | SNPstream panel I | Marker did not work |
| L110PCT_F | AATATTCCACACAAGAAAACGG | - | SNPstream panel I | Monomorphic marker |
| L110PCT_R | RGGGTGGTCTGGAGTCTTC | - | SNPstream panel I | Monomorphic marker |
| L110PCT_U8 | - | GTGATTCTGTACGTGTCGCCGAYGAYTCCGCAGCGTCCAGMTCCC | SNPstream panel I | Monomorphic marker |
| rs2741919_F | TTCAGGAGCACCTGAGAGTG | - | SNPstream panel I | - |
| rs2741919_R | ACTGATGACACACCATGGATT | - | SNPstream panel I | - |
| rs2741919_CT_U8 | - | GTGATTCTGTACGTGTCGCCCACYCACCAGGGGCGGATTATGCAA | SNPstream panel I | - |
| I591T_F | TTCCTTGTCTTTCCTTGTTGTC | - | SNPstream panel I | Below 5% minor allele frequency |
| I591T_R | ACTTGCCTTGATCTGGGC | - | SNPstream panel I | Below 5% minor allele frequency |
| I591T_CT_U9 | - | GACCTGGGTGTCGATACCTATTTTTCTCCGTAGTTCCRGAGCTGA | SNPstream panel I | Below 5% minor allele frequency |
| rs417929_F | AAAGTGGGGTGTGGCAGG | - | SNPstream panel I | Monomorphic marker |
| rs417929_R | CTTTATTTTTTACAAACCCTCTTGTT | - | SNPstream panel I | Monomorphic marker |
| rs417929_CT_U10 | - | AGATAGAGTCGATGCCAGCTTAACCCTGCTGAAGGGCTGCAGTCA | SNPstream panel I | Monomorphic marker |
| rs224243_F | AATATGATTCCTTTCGGATAACTTG | - | SNPstream panel I | - |
| rs224243_R | AGCAACATCCCGGTTGTC | - | SNPstream panel I | - |
| rs224243_CT_U10 | - | AGATAGAGTCGATGCCAGCTGATGGAAAATCACAGGCCAAAGTCA | SNPstream panel I | - |
| rs1149487_F | ATCTTCATGTGTGGGCACC | - | SNPstream panel I | Marker did not work |
| rs1149487_R | ATTAAGAATTGTTTCCAGCTGC | - | SNPstream panel I | Marker did not work |
| rs1149487_CT_U11 | - | AGAGCGAGTGACGCATACTAGCTGCGGCTGGAACCACAGTGGACA | SNPstream panel I | Marker did not work |
| T2671_F | AACTTTAATATCCAAGGGGATTC | - | SNPstream panel I | Below 5% minor allele frequency |
| T2671_R | TTCTCTGCAGCCGATATAAAGTA | - | SNPstream panel I | Below 5% minor allele frequency |
| T2671_CT_U12 | - | CGACTGTAGGTGCGTAACTCATTCTCCTGACTCTAGAGGAAAAGA | SNPstream panel I | Below 5% minor allele frequency |

| | | | | |
|---|---|---|---|---|
| rs401298_F | AAATCTTACCTCAGTCTTGCGA | - | SNPstream panel II | - |
| rs401298_R | TTCTGGAATTGTATTCATCAGGT | - | SNPstream panel II | - |
| rs401298_GA_U1 | - | ACGCACGTCCACGGTGATTTGTCTTACAAACCAAATTGGAAG | SNPstream panel II | - |
| rs224230_F | AAGAATCTGTAGCTTAGTGACTTGC | - | SNPstream panel II | - |
| rs224230_R | TCAGTGGTCAGCTGGAAAT | - | SNPstream panel II | - |
| rs224230_GA_U2 | - | GGATGGCGTTCCGTCCTATTCATGACAATAATTTGTGATTTTACT | SNPstream panel II | - |
| rs8054015_F | TCATGATATGTTTATACAGTGGAAG | - | SNPstream panel II | - |
| rs8054015_R | TTATAACTTGCCTTTTCACTTAAAGTATC | - | SNPstream panel II | - |
| rs8054015_GA_U3 | - | CGTGCCGCTCGTGATAGAATATATGTTTATACAGTGGAAGCCAAC | SNPstream panel II | - |
| rs224217_F | ATATATGCACAGCTTCACAAATGT | - | SNPstream panel II | - |
| rs224217_R | TGTGAGGTTATTGTGAGAGGG | - | SNPstream panel II | - |
| rs224217_GA_U4 | - | AGCGATCTGCGAGACCGTATGTCTTCCCCAGATTTTCTACCTGGT | SNPstream panel II | - |
| rs71999_F | AAACTCCTATGCCAACTCCT | - | SNPstream panel II | - |
| rs71999_R | GCAAGTTTACCTGCTCTCATC | - | SNPstream panel II | - |
| rs71999_GA_U5 | - | GCGGTAGGTTCCCGACATATTGTCTTCACTACTAGAGTGTAGATT | SNPstream panel II | - |
| R408Q_F | CYATCTGCCTCATCTGCAG | - | SNPstream panel II | Below 5% minor allele frequency |
| R408Q_R | TGCCTACCTTGTGTTCCAG | - | SNPstream panel II | Below 5% minor allele frequency |
| R408Q_GA_U6 | - | GGCTATGATTCGCAATGCTTCTGAGTCAGGAGCACCAAGGCCACC | SNPstream panel II | Below 5% minor allele frequency |
| rs224208_F | AAACAAACTGAAGCGCTGAA | - | SNPstream panel II | - |
| rs224208_R | ATCTGGCCCACGTCCTCC | - | SNPstream panel II | - |
| rs224208_GA_U7 | - | AGGGTCTCTACGCTGACGATTGGAGCAGGTGTACTACTTCCTGGA | SNPstream panel II | - |
| rs224212_F | TTTCTGGTAAGGTCAGAGGTG | - | SNPstream panel II | - |
| rs224212_R | AACCACAGCAGAATCTCGG | - | SNPstream panel II | - |
| rs224212_GA_U8 | - | GTGATTCTGTACGTGTCGCCGATGGCCCATCCGTCCCTGGGAGGA | SNPstream panel II | - |
| rs224226_GA_F | TATAATCTGGGTTTTGCTGACC | - | SNPstream panel II | - |

| | | | | |
|---|---|---|---|---|
| rs224226_GA_R | ATTATTTATAAGTGTGAGCCACCAT | - | SNPstream panel II | - |
| rs224226_GA_U8 | - | GTGATTCTGTACGTGTCGCCCAGTTGTCCTGCTTAAAACGTTTGA | SNPstream panel II | - |
| rs224205_F | AGCCCTGGTAAGTGCAGC | - | SNPstream panel II | Marker did not work |
| rs224205_R | AAAAAAGAAGGAAACTGTCGGT | - | SNPstream panel II | Marker did not work |
| rs224205_GA_U9 | - | GACCTGGGTGTCGATACCTAGAGCTAAAAGTCCAGGAGCCCAGAA | SNPstream panel II | Marker did not work |
| rs224231_GA_F | GGGCACGGGACTGATACTC | - | SNPstream panel II | - |
| rs224231_GA_R | ACGGCAAAACCTCGTCTCTA | - | SNPstream panel II | - |
| rs224231_GA_U9 | - | GACCTGGGTGTCGATACCTAAACCTCCACCTCCCAGGTTCAAGCA | SNPstream panel II | - |
| rs1231124_F | GTGGCCCAAGTACCCGTG | - | SNPstream panel II | - |
| rs1231124_R | AAAGTTGAGGACCAGCATTTAG | - | SNPstream panel II | - |
| rs1231124_GA_U10 | - | AGATAGAGTCGATGCCAGCTCTGGAAATGAACTACATTCTCCACA | SNPstream panel II | - |
| rs442387_F | TCTGGTGACTCCCATAGGT | - | SNPstream panel II | - |
| rs442387_R | AAACCGAGACAGTAGGAGAAGG | - | SNPstream panel II | - |
| rs442387_GA_U11 | - | AGAGCGAGTGACGCATACTAGCAGTGTAGAAGTTAGCAAGCTGGA | SNPstream panel II | - |
| M694V_F | GAGGTGGAGGTTGGAGACAA | - | SNPstream panel II | Below 5% minor allele frequency |
| M694V_R | AGAGCAGCTGGCGAATGTAT | - | SNPstream panel II | Below 5% minor allele frequency |
| M694V_GA_U12 | - | CGACTGTAGGTGCGTAACTCGAATGGCTACTGGGTGGTGATAATG | SNPstream panel II | Below 5% minor allele frequency |
| rs224215_GA_F | TCCCAGAAATTCACAGCACA | - | SNPstream panel II | - |
| rs224215_GA_R | GCAAGATGGAGGAACAGAGG | - | SNPstream panel II | - |
| rs224215_GA_U12 | - | CGACTGTAGGTGCGTAACTCTTCTAGAGCTTTCCCGAGGGATACT | SNPstream panel II | - |
| rs1149483_F | GAGCATCTTGTTGCTTACTTGG | - | SNPstream panel III | - |
| rs1149483_R | GCAGTTGACAGCGGGTAAGT | - | SNPstream panel III | - |
| rs1149483_CT_U1 | - | ACGCACGTCCACGGTGATTTGTCTGAAAATAGGCTACTTTTTTAC | SNPstream panel III | - |
| rs11645952_F | TCAGTCGAGCGTCTGTAAGG | - | SNPstream panel III | Below 5% minor allele frequency |
| rs11645952_R | TGTCAGCGCTAAGATTGGTG | - | SNPstream panel III | Below 5% minor allele frequency |

| | | | | |
|---|---|---|---|---|
| rs11645952_CT_U2 | - | GGATGGCGTTCCGTCCTATTTACGATTCACCCAAATCTGATAATG | SNPstream panel III | Below 5% minor allele frequency |
| rs6501170_F | CTCATTGAGAACACAGGAGCTG | - | SNPstream panel III | - |
| rs6501170_R | AGTATCCCTTCCGGACCACT | - | SNPstream panel III | - |
| rs6501170_CT_U3 | - | CGTGCCGCTCGTGATAGAATAACACCAGGCATCCCAGATGCCTCG | SNPstream panel III | - |
| rs6501169_F | TCTCAGCTGTGCCGTCGT | - | SNPstream panel III | - |
| rs6501169_R | TGACAGGCCTGATTCATAACA | - | SNPstream panel III | - |
| rs6501169_CT_U4 | - | AGCGATCTGCGAGACCGTATAAACCAAAAGGAAAGGCTCAAGGAA | SNPstream panel III | - |
| rs8052682_F | CCACTCTTGCTCTAGTACCTTGC | - | SNPstream panel III | - |
| rs8052682_R | TGTCTGGAGGAGGCTTGAAT | - | SNPstream panel III | - |
| rs8052682_CT_U5 | - | GCGGTAGGTTCCCGACATATTGCTGCGATTACGGGTGTGAGCCAC | SNPstream panel III | - |
| rs250470_F | ACTTGCCTGGTGGAGAACAG | - | SNPstream panel III | - |
| rs250470_R | CTCCGTCAGTGAGCCAAGAT | - | SNPstream panel III | - |
| rs250470_CT_U6 | - | GGCTATGATTCGCAATGCTTCAACACATGAAATAAAACAGTTGGG | SNPstream panel III | - |
| rs458522_F | GGCGCTTACTTCTCTTCACTACTT | - | SNPstream panel III | - |
| rs458522_R | CCAATGCAGTTCCAGAGGAT | - | SNPstream panel III | - |
| rs458522_CT_U7 | - | AGGGTCTCTACGCTGACGATAAGTAGCAGTTCTTAGTCTTTTTGG | SNPstream panel III | - |
| rs220380_F | AGAGGAGCCTGTTGGCAAG | - | SNPstream panel III | Marker did not work |
| rs220380_R | GACAGTCAGGTGCCCAAATC | - | SNPstream panel III | Marker did not work |
| rs220380_CT_U8 | - | GTGATTCTGTACGTGTCGCCTCCTACCCTCTTGGTCTTTTTAAAG | SNPstream panel III | Marker did not work |
| rs190081_F | AATGTCTCCATGCTCTACAGGTT | - | SNPstream panel III | - |
| rs190081_R | CAGTTGTTAAGAGTGGCAATGG | - | SNPstream panel III | - |
| rs190081_CT_U9 | - | GACCTGGGTGTCGATACCTATGACATCTATCTAGGTCCAGATGAA | SNPstream panel III | - |
| rs188760_F | ACATGCAGTATTCCAGGATGC | - | SNPstream panel III | - |
| rs188760_R | TGTGAGAGATGATAGGCTGGAA | - | SNPstream panel III | - |
| rs188760_U10 | - | AGATAGAGTCGATGCCAGCTTTCCGTTAGTGTTTCTAGGCTCGTC | SNPstream panel III | - |

| | | | | |
|---|---|---|---|---|
| rs7500738_F | AAGAGACTGCTTCAGGCATCA | - | SNPstream panel III | - |
| rs7500738_R | CGTGCTCAGGTGTCACTGTC | - | SNPstream panel III | - |
| rs7500738_CT_U11 | - | AGAGCGAGTGACGCATACTATCTGTACTTAAAAAGTCAGGATCAG | SNPstream panel III | - |
| rs757108_F | CAAGAACAGCTCAGCAGGAA | - | SNPstream panel III | - |
| rs757108_R | CTACTGCTGCCTGGAAGGAC | - | SNPstream panel III | - |
| rs757108_CT_U12 | - | CGACTGTAGGTGCGTAACTCAATAGTTTATTCACCCAAAATGTCA | SNPstream panel III | - |
| rs224230b_F | TGTGCAGCCAAGAATCTGTA | - | SNPstream panel IV | - |
| rs224230b_R | CTCAGTGGTCAGCTGGAAAT | - | SNPstream panel IV | - |
| rs224230_GA_U1 | - | ACGCACGTCCACGGTGATTTCATGACAATAATTTGTGATTTTACT | SNPstream panel IV | - |
| rs8054015b_F | AAGATTCTGCCACTGCACTC | - | SNPstream panel IV | - |
| rs8054015b_R | CTGCTTTATAACTTGCCTTTTCA | - | SNPstream panel IV | - |
| rs8054015_GA_U2 | - | GGATGGCGTTCCGTCCTATTATATGTTTATACAGTGGAAGCCAAC | SNPstream panel IV | - |
| rs250470b_F | GTACTCCTGCCTGGGTGAC | - | SNPstream panel IV | - |
| rs250470b_R | CTAAACTTGCCTGGTGGAGA | - | SNPstream panel IV | - |
| rs250470_GA_U3 | - | CGTGCCGCTCGTGATAGAATATTAATTAATTAATTTTCCTAGTCC | SNPstream panel IV | - |
| rs224208b_F | ATACCTCCCTGTCCTCTGCT | - | SNPstream panel IV | - |
| rs224208b_R | CATATGCCTTCCTGATCTGC | - | SNPstream panel IV | - |
| rs224208_GA_U4 | - | AGCGATCTGCGAGACCGTATTGGAGCAGGTGTACTACTTCCTGGA | SNPstream panel IV | - |
| rs224243b_F | TCCATGTTCCAGGACAGACT | - | SNPstream panel IV | - |
| rs224243b_R | CACGGAAATTCAGAAACACC | - | SNPstream panel IV | - |
| rs224243_GA_U5 | - | GCGGTAGGTTCCCGACATATGTCATGGGTGACAGCCTTCCTTCAG | SNPstream panel IV | - |
| rs224215b_F | AGCCTCCTTTGCAGTTAGGT | - | SNPstream panel IV | - |
| rs224215b_R | CCTCAAGCTTCTGCTCTCAG | - | SNPstream panel IV | - |
| rs224215_GA_U6 | - | GGCTATGATTCGCAATGCTTTTCTAGAGCTTTCCCRAGGGATACT | SNPstream panel IV | - |
| rs224241b_F | GCTCCAGAGCAGAAACTGAA | - | SNPstream panel IV | - |

| | | | | |
|---|---|---|---|---|
| rs224241b_R | TGCCCTCTACAGGACAGAAA | - | SNPstream panel IV | - |
| rs224241_GA_U7 | - | AGGGTCTCTACGCTGACGATGATAAACTGAAATGACTGACAGGCC | SNPstream panel IV | - |
| rs6501170b_F | TTAACCGATGCCACTGAACT | - | SNPstream panel IV | - |
| rs6501170b_R | CCTCCCTGGAGTATCCATTT | - | SNPstream panel IV | - |
| rs6501170_GA_U8 | - | GTGATTCTGTACGTGTCGCCGTATCCCTTCCGGACCACTCTCATT | SNPstream panel IV | - |
| rs190081b_F | TGTACCCAAATATGCACGTCT | - | SNPstream panel IV | - |
| rs190081b_R | TTAGCAGTGTGCCTTTGACA | - | SNPstream panel IV | - |
| rs190081_GA_U9 | - | GACCTGGGTGTCGATACCTACAGTTGTTAAGAGTGGCAATGGCCT | SNPstream panel IV | - |
| rs224217b_F | GAATCTCAACCCCATGATGA | - | SNPstream panel IV | - |
| rs224217b_R | TGCACAGCTTCACAAATGTC | - | SNPstream panel IV | - |
| rs224217_GA_U10 | - | AGATAGAGTCGATGCCAGCTGTCTTCCCCAGATTTTCTACCTGGT | SNPstream panel IV | - |
| rs442387b_F | GGATACAACGCACAAACACA | - | SNPstream panel IV | - |
| rs442387b_R | CTCTGGTGACTCCCATAGGTT | - | SNPstream panel IV | - |
| rs442387_GA_U11 | - | AGAGCGAGTGACGCATACTAGCAGTGTAGAAGTTAGCAAGCTGGA | SNPstream panel IV | - |
| rs1231124b_F | ACTTTTCAGGGACAGGCACT | - | SNPstream panel IV | - |
| rs1231124b_R | CCTCTGAATCCAGGGAAGAC | - | SNPstream panel IV | - |
| rs1231124_GA_U12 | - | CGACTGTAGGTGCGTAACTCCTGGAAATGAACTACATTCTCCACA | SNPstream panel IV | - |
| rs250470_F | ACTTGCCTGGTGGAGAACAG | - | Fluorescent Polarization | - |
| rs250470c_R | ATCCCTTCCTGATGCCTTCT | - | Fluorescent Polarization | - |
| rs2504070_CT_Sense | - | ACACATGAAATAAAACAGTTGGG | Fluorescent Polarization | - |
| rs2504070_GA_AS | - | TTAATTAATTAATTAATTTTCCTAGTCC | Fluorescent Polarization | - |
| R26Q_F | CACTGAATTTGGAGGAGAAAATG | - | Fluorescent Polarization | Below 5% minor allele frequency |
| R26Q_R | ACCACATGTAGACCGCATCA | - | Fluorescent Polarization | Below 5% minor allele frequency |
| R26Q_GA_Sense | - | GTTCTAAGTAGAATTGAAGCTGGAC | Fluorescent Polarization | Below 5% minor allele frequency |
| R26Q_CT_AS | - | TTCATCCATTACACTTTTAAGTGAT | Fluorescent Polarization | Below 5% minor allele frequency |

| | | | | |
|---|---|---|---|---|
| V534I_F | TACAAGTGCCCTGAGTGTGG | - | Fluorescent Polarization | Below 5% minor allele frequency |
| V534I_R | GGGCTCCATGGTTTGTAAGA | - | Fluorescent Polarization | Below 5% minor allele frequency |
| V534I_GA_Sense | - | TTCTCTCGGAGTTCACACCTC | Fluorescent Polarization | Below 5% minor allele frequency |
| V534I_CT_AS | - | CTCATGAGTTCTTTCGTGAATGA | Fluorescent Polarization | Below 5% minor allele frequency |
| C310S_F | TTCTGGGCTGCTGACCTAAC | - | Fluorescent Polarization | - |
| C310S_R | TCATGAGGTCTTCCCAGCTC | - | Fluorescent Polarization | - |
| C310S_GC_Sense | - | TGGAAGGTGTTCCGTCTGTAT | Fluorescent Polarization | - |
| C310S_CG_AS | - | TGAGGGTGGATGTTCTCAGAG | Fluorescent Polarization | - |
| N248H_F | TAAGGCGCGACAGACACTC | - | Fluorescent Polarization | Below 5% minor allele frequency |
| N248H_R | CTGCCAATCAGACCAACAGA | - | Fluorescent Polarization | Below 5% minor allele frequency |
| N248H_AC_Sense | - | CGGGAGCACGCGAGCC | Fluorescent Polarization | Below 5% minor allele frequency |
| N248H_TG_AS | - | GCTAATCCGCCCTCCAAT | Fluorescent Polarization | Below 5% minor allele frequency |
| rs224230c_F | CACAACATCGTCTCCTGCAT | - | Fluorescent Polarization | - |
| rs224230c_R | GAGCTGGGTGTTCAGCTAGG | - | Fluorescent Polarization | - |
| rs224230_GA_Sense | - | CATGACAATAATTTGTGATTTTACT | Fluorescent Polarization | - |
| rs224230_CT_AS | - | TGACATCCATGGGTAGTCATTAA | Fluorescent Polarization | - |
| E148Q_F | AAAACGGCACAGATGATTCC | - | Fluorescent Polarization | Below 5% minor allele frequency |
| E148Q_R | TCTCTCTGCGTTTGCTCAGG | - | Fluorescent Polarization | Below 5% minor allele frequency |
| E148Q_GC_Sense | - | TGCGGTGCAGCCAGCCC | Fluorescent Polarization | Below 5% minor allele frequency |
| E148Q_CG_AS | - | CAGCCCCCTCCCGGCCT | Fluorescent Polarization | Below 5% minor allele frequency |
| rs182674_F | TCCGATTGAAAAAGGCAATC | - | Fluorescent Polarization | - |
| rs182674_R | AAAGAAGATGGCCAGGGTTT | - | Fluorescent Polarization | - |
| rs182674_GA_Sense | - | AGGCGCTGTCCCCAGC | Fluorescent Polarization | - |
| rs182674_CT_AS | - | ATGCGATGTAGGAAGAAGCA | Fluorescent Polarization | - |
| rs224217c_F | CAAATTTCAGGTAGTTCTTGGACAC | - | Fluorescent Polarization | - |

| | | | | |
|---|---|---|---|---|
| rs224217c_R | GATGATGCATTGCTGTGAGG | - | Fluorescent Polarization | - |
| rs224217_CT_Sense | - | GAATCCAAGGGGTTTATGGA | Fluorescent Polarization | - |
| rs224217_GA_AS | - | CCCCAGATTTTCTACCTGGT | Fluorescent Polarization | - |
| rs224213_F | TCTGTGTAAGCAACTTGGGTTTG | - | Fluorescent Polarization | - |
| rs224213_R | ACAGGTACCGTCAACTGGGTCT | - | Fluorescent Polarization | - |
| rs224213_CT_Sense | - | GGCTGCGAGTCCCCG | Fluorescent Polarization | - |
| rs224213_GA_AS | - | TTCCTGGGCGTGGCA | Fluorescent Polarization | - |
| rs224208c_F | AGCCCACCTCTTATCCACCT | - | Fluorescent Polarization | - |
| rs224208c_R | CACCTGCAGAAGTTCCCATT | - | Fluorescent Polarization | - |
| rs224208_GA_Sense | - | AGCAGGTGTACTACTTCCTGGA | Fluorescent Polarization | - |
| rs224208_CT_AS | - | CACAAAGAAATGCTCYTGCTG | Fluorescent Polarization | - |
| rs224205b_F | TGCTCCACTTCCACTGACAC | - | Fluorescent Polarization | Marker did not work |
| rs224205b_R | TCTCCCCCATATGCTTTCTG | - | Fluorescent Polarization | Marker did not work |
| rs224205_GA_Sense | - | AAGTCCAGGAGCCCAGAA | Fluorescent Polarization | Marker did not work |
| rs224205_CT_AS | - | GGCTGACTCCTGGCCTCTA | Fluorescent Polarization | Marker did not work |
| rs8054015_GA_Sense | - | ATATGTTTATACAGTGGAAGCCAAC | Fluorescent Polarization | - |
| rs8054015_CT_AS | - | TTAAAGTATCTCTTTTCAATGGTTC | Fluorescent Polarization | - |
| rs401298b_F | CGGGTCCAAATCTTACCTCA | - | Fluorescent Polarization | - |
| rs401298b_R | AGACCCAATCCTCACCCTCT | - | Fluorescent Polarization | - |
| rs401298_GA_Sense | - | TGTCTTACAAACCAAATTGGAAG | Fluorescent Polarization | - |
| rs401298_CT_AS | - | GATGTTATATTCTTGCCTTTTGTAA | Fluorescent Polarization | - |

**Supplementary Table 3.3: List of oligos used to amplify fragments for the sequencing experiments**

| Oligo ID | Sequence |
|---|---|
| MEFVprom1F | GGGTTTGGGCTCAGAAAGAT |
| MEFVprom1R | CCAAACGTCAAACCACTTCA |
| MEFVProm2_F | TCAGTGGTCAGCTGGAAATG |
| MEFVProm2_R | GGCAGGAGAATCACTTGAACA |
| MEFVprom3F | CCCAAGGTGGAGTACAATGG |
| MEFVprom3R | CAAGTCTGCAAGGGAAGGTC |
| MEFVexon1F | GGCAGGAAGGAAGATTGGAG |
| MEFVexon1R | AGCTGCTCTGAGCTCCTGGT |
| MEFVexon2F | GGGGATTCTCTCTCCTCTGC |
| MEFVexon2R | GATTACAGGCATGAGCTATCG |
| MEFVexon3F | GGGGAGGACAAGCTAGGAAG |
| MEFVexon3R | TAATGCACCAACAACCCAGA |
| MEFVexon4F | CCTCAGCCTTGCTACCAGAA |
| MEFVexon4R | TCTGTCCCCTGAGAGGAGGT |
| MEFVexon5F | CTGGGGGTTCCTGGACAT |
| MEFVexon5R | GGTCACCAAGACCAAGTCCT |
| MEFVexon6F | GCCTGAATTCCCGTGGTTAG |
| MEFVexon6R | GAACATCTCCCTCCCAGGTC |
| MEFVexon7-8F | TGTAGTTCATTTCCAGCTCACG |
| MEFVexon7-8R | TTCTAAATAGGGCCCCTCAA |
| MEFVexon9-10aF | CTAAGCAGGGGGTTCCTTGT |
| MEFVexon9-10aR | AATAAAGGAGCCTCCCAAGC |
| MEFVexon9-10bF | AGAATGGCTACTGGGTGGTG |
| MEFVexon9-10bR | AATCCAGTCTGCTTGCGTTT |
| MEFVexon9-10cF | AACAGAAGATTTGGCCCTCA |
| MEFVexon9-10cR | GTGGCTCATGCCTGTAATCC |
| MEFVexon9-10dF | GTAGAGACGGGGGTTTCACC |
| MEFVexon9-10dR | CATGAGGCCCAGTCAATTCT |
| AK096958prom1_F | CTGAGATGAGGGGACACTGG |
| AK096958prom1_R | GTGGGTTGGTCCAAGTTGAG |
| AK096958prom2F | TTGCTGAGTGTGGTGGTGAT |
| AK096958prom2R | TGGTGTGGGGATCTATTCAGA |
| AK096958prom3F | CTCCTGAGCAGCTTGGAAAC |
| AK096958prom3R | GCCTGTAATCCCACCACTTT |
| AK096958exon1aF | GTCAGGCTGGTCTCGAACTC |
| AK096958exon1aR | ACACTGCGGACAGTTCCTTT |

| | |
|---|---|
| AK096958exon1bF | GGAGGAGACGTGTGGAAAGA |
| AK096958exon1bR | TCTCAGCCTCCTTGGAACAG |
| AK096958exon1cF | CTGTGCTGACCCCAAAACC |
| AK096958exon1cR | GTCAGGAGAGCGAGACCATC |
| AK096958exon1dF | CCTTCCGAGTAGCTGGGATT |
| AK096958exon1dR | AGTGCCCTGGTGACAGACAT |
| AK096958exon2F | ATGACTGACAGGCCATTTGC |
| AK096958exon2R | GAACCCCCATTTTACAGTTGA |
| ZNF263prom1F | TTCAGGAACTTGCTGTGCAT |
| ZNF263prom1R | CTGCCTCAGCCTCCTGAGTA |
| ZNF263prom2F | TGGTGAAACCCCGTCTCTAC |
| ZNF263prom2R | TTTAAATCTGCCCGAGATGTG |
| ZNF263prom3F | CCCCCAACTATCACAGTCCA |
| ZNF263prom3R | TGAAATTCTGGGAGTGACTGG |
| ZNF263exon1aF | GACACTGACGCACTGGAGAG |
| ZNF263exon1aR | CAGTCCTCCTCCAGCTTCAC |
| ZNF263exon1bF | TCCAACCTTACATGGGTTCA |
| ZNF263exon1bR | CAAGTCCTCTTCCCCACGTA |
| ZNF263exon2F | GATCGCCTGAGGTTCTCTGT |
| ZNF263exon2R | CACCAACTGCAGGTCTTTCA |
| ZNF263exon3-4F | AGCTGATTAGGCCTCTGTGC |
| ZNF263exon3-4R | AGGCTTCTTGGCAAAACTCA |
| ZNF263exon5F | TCCACTGGGAACAGTTTGTG |
| ZNF263exon5R | TGCCTCACTCTTAACTGCAAGA |
| ZNF263exon6aF | TTCTGGGCTGCTGACCTAAC |
| ZNF263exon6aR | CCCACACTCGGGACACTTAT |
| ZNF263exon6bF | GCACCAGAGAACGCACACT |
| ZNF263exon6bR | CTCTGCACCTCCCTCACAAT |
| ZNF263exon6cF | TGAGGTGGCATATTCAGAGG |
| ZNF263exon6cR | GATCAACACTCGGGGGTCTA |
| TIGD7prom1F | CAGTAGCGCCATCAGCATTA |
| TIGD7prom1R | CAGGCCACTATGACAGTTGTGT |
| TIGD7prom2F | AAAAAGACGACGGCACAGC |
| TIGD7prom2R | GCCATCTCGGCTCACTACA |
| TIGD7prom3F | AAAAATTAGCCGGGTGTGGT |
| TIGD7prom3R | TTCTCAGAGATGAGCCCAAAA |
| TIGD7exon1aF | CACTGAAGAGAAATCAAAATCTCAA |
| TIGD7exon1aR | CAAGATGTGCCCACAAGACA |
| TIGD7exon1bF | CTTTTTGGAAGGGGAGAAGC |
| TIGD7exon1bR | CTTACTGATTCCAAATTCATCCA |
| TIGD7exon1cF | TCTCCTGTTGGTGGTTCAGA |
| TIGD7exon1cR | TTGGGCAGTTTTGATTTTCC |
| TIGD7exon1dF | CAGTGGGGATGAAACAGACC |

| | |
|---|---|
| TIGD7exon1dR | TCCCCATGTTCTAAGCCTTG |
| TIGD7exon1eF | GGCAAAAAGTTGGGAAGAAG |
| TIGD7exon1eR | GCAAAGGCATTCCACCTTAG |
| ZNF75Aprom1F | TGCTAGCAGAACTTCACTTATTCAA |
| ZNF75Aprom1R | CGCCTGTAGTCCCAGCTACT |
| ZNF75Aprom2F | TTTTGCGACGGAGTCTCG |
| ZNF75Aprom2R | TGTAGCCAATATTGCAGAGTTGA |
| ZNF75Aprom3F | CCCAACCCCAAATCAACTT |
| ZNF75Aprom3R | TCGAGGAAGAGGCATTTTGT |
| ZNF75Aexon1F | TGCGTCACGAGACCTAGAAA |
| ZNF75Aexon1R | CACCGCCTACGGAAAATG |
| ZNF75Aexon2F | CAATGTGGGCTGTACACCTG |
| ZNF75Aexon2R | TTGCTATCAAGGCTGGTTTT |
| ZNF75Aexon3F | CTAGTTCTAGGGGCTGGTGGA |
| ZNF75Aexon3R | AGCCTGGGAGGTACCCAGTT |
| ZNF75Aexon4F | CTTTATCGCCTGGCCAATTC |
| ZNF75Aexon4R | CCTACACCCTCCCCTATACCA |
| ZNF75Aexon5F | TCCATTTAACCATTTGGAGTAACA |
| ZNF75Aexon5R | GCCCTTGCAAAAGTTCACAT |
| ZNF75Aexon6aF | CCTGTGATTTGCTTGTTCTGG |
| ZNF75Aexon6aR | CCAAGCTCCATTTGGTGAAT |
| ZNF75Aexon6bF | AGACACCAGAAACTCCACCTG |
| ZNF75Aexon6bR | TGGGCAAGTACTAATTTTTATAGCTG |
| ZNF75Aalt1aF | TCATAAGTCGCAAGTGATGAGG |
| ZNF75Aalt1aR | AAGCCAGACTAGGCCCAAC |
| ZNF75Aalt1bF | GGACAGTTAAGGTGGGACTTGA |
| ZNF75Aalt1bR | CTTGGCCTCTCAAAGTGCTG |

# 4 INVESTIGATION OF *NLRP3-MEFV* GENE-GENE INTERACTIONS IN CROHN'S DISEASE AND ULCERATIVE COLITIS

*The following data have not been sent for publication and were only put together for the purpose of this Ph.D. thesis.*

## 4.1 Rationale

We decided to pursue the following analyses based on the results from Chapters 2 and 3. Since we knew that the gene product of *NLRP3* and *MEFV* interacted with each other, we performed additional gene-gene interaction analyses to assess whether such interactions could contribute to CD and/or UC susceptibility.

# 4.2 *NLRP3-MEFV* gene-gene interactions do not contribute to inflammatory bowel disease susceptibility

## 4.2.1 Summary

As reported in Chapter 3, none of the significant associations observed in the *MEFV* region were validated across the replication sample sets. Given that we have previously identified 6 common SNPs in the *NLRP3* region contributing to CD susceptibility (Villani et al. 2009; Chapter 2), that *NLRP3* and *MEFV* gene products (NALP3 and pyrin respectively) are known to interact together in the inflammasome molecular platform and to be involved in similar pathways (Chae et al. 2006; Papin et al 2007; Chae et al. 2008), we subsequently evaluated whether possible gene-gene interaction between the *MEFV* tagging SNPs (16 in total; Chapter 3) and the *NLRP3* SNPs (6 in total; Chapter 2) could have masked a contribution of *MEFV* to CD susceptibility. In addition, we also assessed the association of the 6 *NLRP3* SNPs (Chapter 2) in the Belgian and Canadian UC sample sets used in the *MEFV* study (Chapter 3). We subsequently evaluated whether gene-gene interactions in the combined Belgian-Canadian UC sample set, commonly genotyped in both studies (Chapters 2-3), could explain the flip-flop phenomenon observed with the R202Q (rs224222) variant in the *MEFV* study (Chapter 3), where different alleles were associated with UC in the Belgian and Canadian samples.

### 4.2.2 Methods

**Genotyping.** Samples were genotyped using Sequenom homogenous MassExtend (hME) assays (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry) (van den Boom et al. 2007), using the primers and probes described in the **Supplementary Table 2.1**.

**Statistical analysis.** Tests of association were performed using the likelihood methods implemented in UNPHASED v3.0.10 (Dudbridge 2008), which can analyze samples of nuclear families, unrelated subjects or a combination of both. It also allows tests to be conditional on each individual's genotype at one marker to test for association at another, to look for associations that may have been missed due to the possibility of epistasis between two loci. It also tests for the presence of gene-gene interaction (merely, the departure from log-additive effects of two or more SNPs on the risk).

### 4.2.3 Results

Using 710 CD trios, 239 CD cases, and 107 healthy controls commonly genotyped in the *NLRP3* (Chapter 2) and *MEFV* study (Chapter 3), stratified analyses of *MEFV* variants conditional on each individual's genotypes at *NLRP3* variants did not provide evidence of associations under the possibility of epistasis (**Figure 4.1**).

No significant associations were observed between the 6 *NLRP3* SNPs and UC in the Belgian, Canadian, and combined Belgian-Canadian sample sets (**Table 4.1**). Moreover no significant interactions were found between R202Q variant (rs224222) located in *MEFV* exon 2 and the 6 SNPs in *NLRP3*. Hence, multi-locus effects between the two genes do not explain the observed flip-flop at R202Q (see Discussion). Additionally, the levels of association for the 16 *MEFV* SNPs conditional on each individual's genotypes at *NLPR3* are in the same order of magnitude than the unconditional levels of association, indicating that epistasis between the two genes is not a mechanism that would be a regulator of the risk of *MEFV* on UC (**Figure 4.2**).

### 4.2.4 Figures

**Figure 4.1: Conditional tests of association and tests of gene-gene interactions between SNPs in *MEFV* and in *NLRP3* in the combined Belgian-Canadian CD sample set**

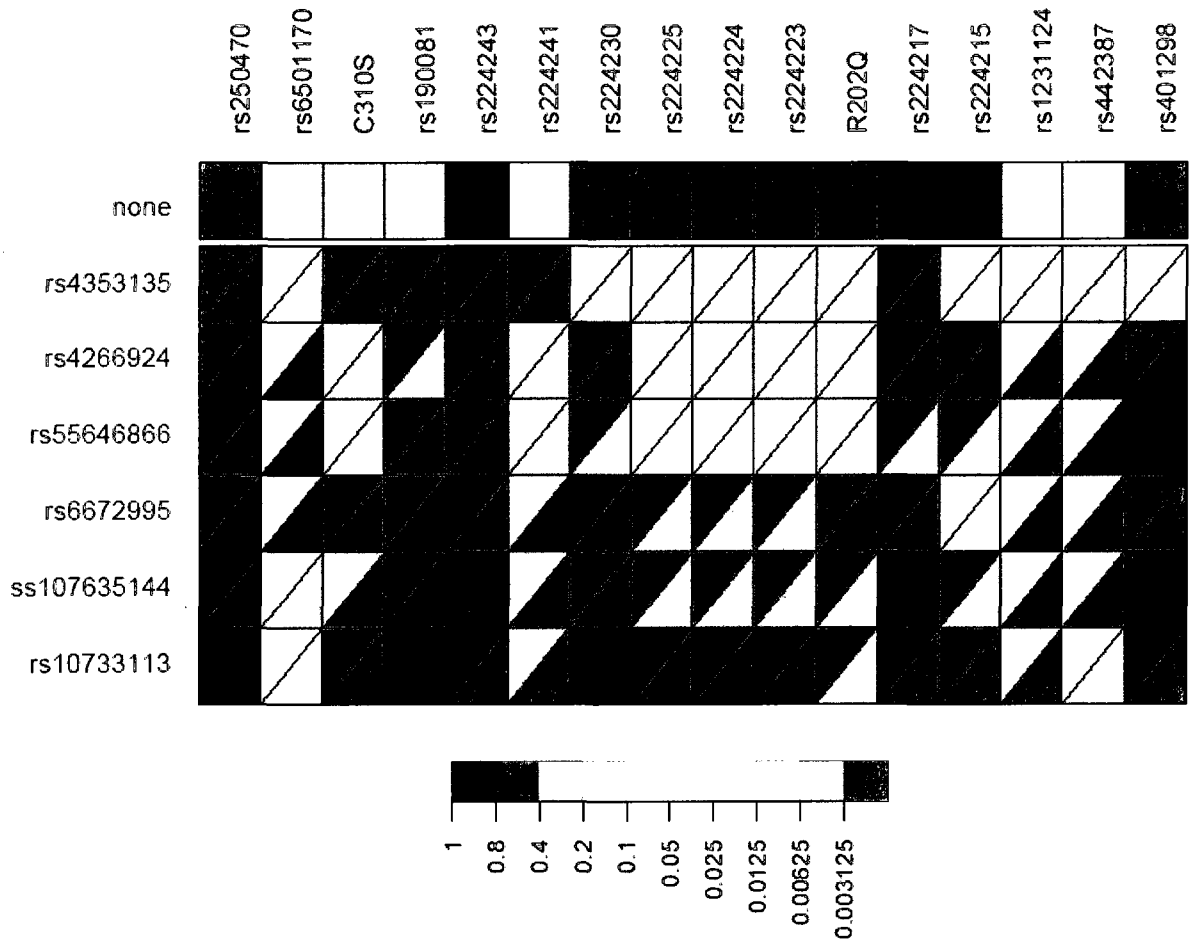



**Figure 4.1:** *P*-values for the tests are color coded and represented. For each combination of SNPs, the upper triangular portions of each squares represent the *p*-value for testing the association of the *MEFV* SNP (horizontal axis) conditional on each individual's genotype at the *NLRP3* SNP (vertical axis). The lower triangular portions of each square represent the *p*-value for the test of statistical interaction between the two SNPs. The unconditional tests of association for each SNP in *MEFV* are shown on the line labeled "none".

**Figure 4.2: Conditional tests of association and tests of gene-gene interactions between SNPs in *MEFV* and in *NLRP3* in the combined Belgian-Canadian UC sample set**

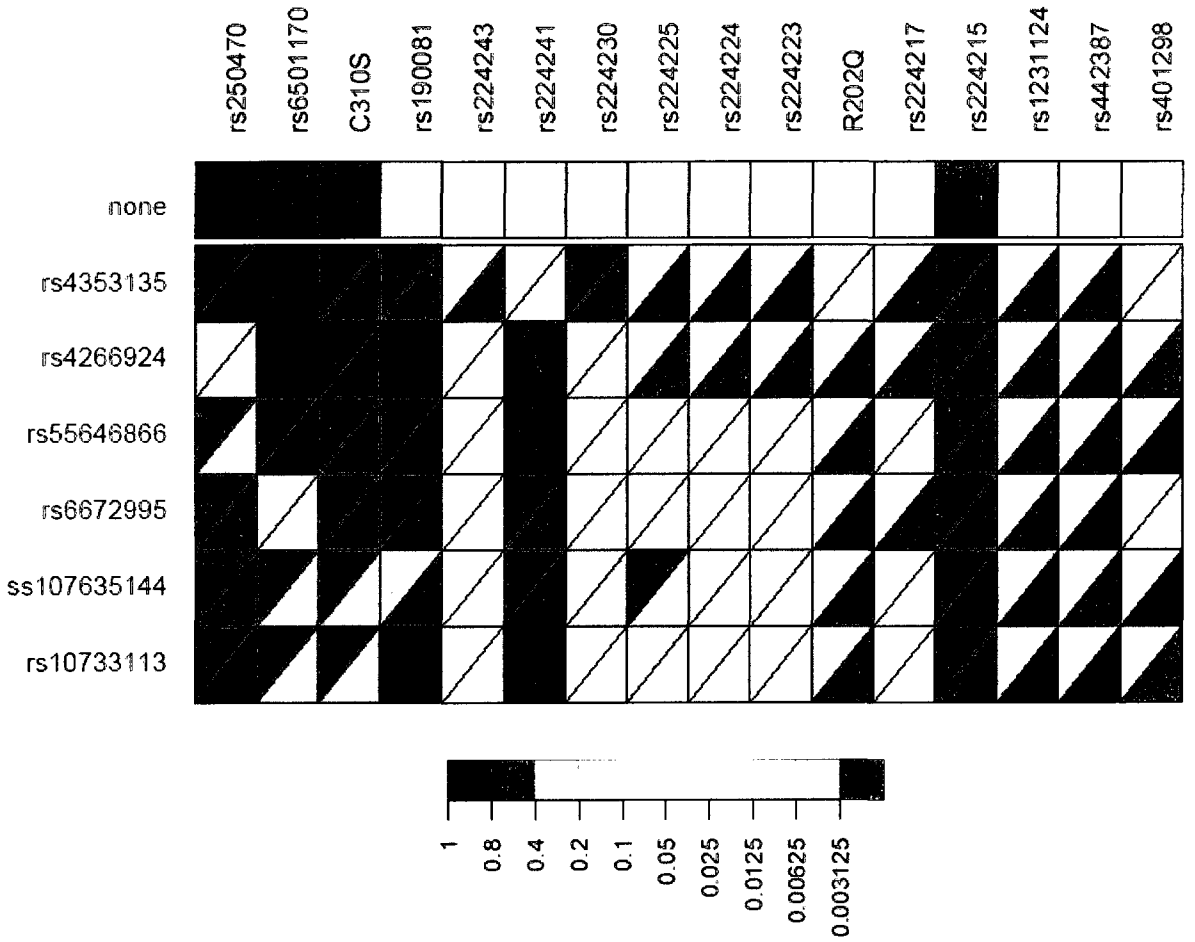



**Figure 4.2:** Refer to legend of **Figure 4.1**.

## 4.2.5 Table

**Table 4.1: Association results of *NLRP3* tagging SNPs in UC sample sets**

| # | SNP | Combined Belgian UC | | | | Combined Canadian UC | | | | Combined UC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Allele | Frequency Cases | Frequency Controls | *P* value | Allele | Frequency Cases | Frequency Controls | *P* value | Allele | Frequency Cases | Frequency Controls | *P* value |
| 2 | rs4353135 | 4 | 0.67 | 0.66 | 0.7506 | 3 | 0.34 | 0.34 | 0.8907 | 4 | 0.67 | 0.66 | 0.7654 |
| 3 | rs4266924 | 1 | 0.87 | 0.81 | 0.3767 | 1 | 0.89 | 0.86 | 0.3345 | 1 | 0.88 | 0.82 | 0.3806 |
| 6 | rs55646866 | 2 | 0.89 | 0.83 | 0.4734 | 2 | 0.90 | 0.85 | 0.2182 | 2 | 0.89 | 0.84 | 0.3046 |
| 8 | rs6672995 | 3 | 0.82 | 0.79 | 0.9175 | 1 | 0.17 | 0.17 | 1.0000 | 3 | 0.83 | 0.80 | 0.9426 |
| 9 | ss107635144 | 2 | 0.85 | 0.79 | 0.3470 | 2 | 0.86 | 0.83 | 0.4645 | 2 | 0.85 | 0.80 | 0.3466 |
| 10 | rs10733113 | 3 | 0.85 | 0.80 | 0.5143 | 3 | 0.86 | 0.83 | 0.6013 | 3 | 0.85 | 0.81 | 0.5195 |

### 4.2.6 Discussion

Since no interactions were observed when looking at the large Belgian-Canadian CD sample set and no common variants were consistently associated with CD in the *MEFV* region (Chapter 3), we concluded that common variants in the *MEFV* region do not contribute to CD susceptibility.

As reported in Chapter 3, only the coding SNP R202Q was commonly associated with UC in the Belgian and Canadian sample set. Yet, we observed opposite allele association with UC in the Belgian and Canadian cohorts, suggesting that R202Q was unlikely to be a causative variant. Such flip-flop association in samples of similar ethnical origins are often regarded as spurious findings. However, different reasons may explain such phenomenon, including a difference in genetic background, ethnicity, and environment (Lin et al. 2007; Zaykin et al. 2008). For example, when attempting to replicate the association of a non-causal allele in LD with the causative variant in two different populations, as it is the case with our Belgian and Canadian sample sets, a difference in LD patterns between the populations could result in inconsistent associations observed across sample sets (Lin et al. 2007; Zaykin et al. 2008). This is not our case, as the Belgian and Canadian UC sample sets displayed very similar LD patterns in the *MEFV* region (data not shown). Such inverse association could also indicate the presence of interactions with another risk locus (Lin et al. 2007). Complex traits usually result from an interplay of several genetic risk loci and environmental factors. Lin *et al.* have shown that performing single marker analysis without considering the possibility of other genetic risk loci or

environmental risk factors correlating with the candidate locus or the possibility of a multi-locus effect could also lead to flip-flop associations (Lin et al. 2007). To assess this latter possibility, since we knew that the *MEFV* and the *NLRP3* gene product interacted with each other, we genotyped the 6 SNPs in the *NLRP3* region that had been previously associated with CD (Villani et al. 2009) in both the Belgian and Canadian UC sample sets and performed gene-gene interaction analysis using these 6 *NLRP3* SNPs and the 16 SNPs (i.e. 12 tagging SNPs and 4 coding SNPs in *MEFV* exon 2 region) genotyped in the *MEFV* region. Since no significant interactions were observed between *NLRP3* variants and R202Q, the inverse allele associations observed with R202Q is still unresolved.

This is also the first report evaluating the 6 *NLRP3* variants associated with CD (Villani et al. 2009; Chapter 2) in a UC sample set. Interestingly, no significant associations between the *NLRP3* variants and UC were observed, similarly to what has been reported about *NOD2* mutations, which are only associated with CD (Cho 2008). Given the similarities between *NOD2* and *NLRP3*, which are both members of the CATERPILLER family (Ting et al. 2006), and the above results, we conclude that the *NLRP3* association is likely to be specific to CD. Both gene products of *NOD2* and *NLRP3* play central roles in intracellular bacterial sensing, and the CD associated variants of both genes were reported to result in a "loss-of-function phenotype" characterized by a decrease in secretion of IL-1$\beta$ (Li et al. 2004; van Heel et al. 2005; Kramer et al. 2006; van Beelen et al. 2007; Villani et al. 2009). Interestingly, the autophagy genes *IRGM* and *ATG16L1* have also been associated specifically to CD (Fisher et al. 2008; Franke et al. 2008b). Together, these results suggest that alterations in the

intracellular sensing and processing of bacteria may constitute a central feature specific to the pathogenesis of CD.

Overall, as we observed no associations between common variants in the *MEFV* region (Chapter 3) and no epistatic interaction between *MEFV* and *NLRP3* variants (Chapter 4), we thus conclude that common variants in *MEFV* do not contribute to CD susceptibility and that the two key regulators of the inflammasomes, *NLRP3* and *MEFV*, do not contribute to UC susceptibility.

# 5 ASSESSMENT OF GENETIC RISK FACTORS FOR POST-INFECTIOUS IRRITABLE BOWEL SYNDROME

The manuscript has been submitted to the *Journal of the American Medical Association*

Reference:

Villani AC, Lemire M, Thabane M, Geneau G, Belisle A, Garg AX, Clark WF, Moayyedi P, Collins SM, Franchimont D, Marshall JK. Genetic risk factors for post-infectious irritable bowel syndrome following a waterborne outbreak of gastroenteritis. *JAMA* (submitted).

**Affiliation of co-authors:**

Alexandra-Chloé Villani[1,2], Mathieu Lemire[3], Marroon Thabane[4], Geneviève Geneau[2], Alexandre Belisle[2], Amit X. Garg[5], William F. Clark[5], Paul Moayyedi[4], Stephen M. Collins[4], Denis Franchimont[1,6-7] and John K. Marshall[4,7]

[1] Division of Gastroenterology, Department of Medicine, McGill University, Montreal, Quebec, Canada; [2] McGill University & Genome Quebec Innovation Centre, Montreal, Quebec, Canada; [3] Ontario Institute for Cancer Research, Toronto, Ontario, Canada; [4] Farncombe Family Digestive Health Research Institute and Division of Gastroenterology, Department of Medicine, McMaster University, Hamilton, Ontario, Canada; [5] Division of Nephrology, Department of Medicine, University of Western Ontario, London, Ontario, Canada; [6] Department of Gastroenterology, Erasme University Hospital, Brussels, Belgium; [7] Drs. Franchimont and Marshall share the senior authorship position

## 5.1 Rationale

For the third candidate gene association study presented in this thesis, we had the opportunity of conducting the first ever published genetic association study in the field of post-infectious irritable bowel syndrome (PI-IBS), using the well-characterized cohort of PI-IBS patients from the Walkerton community that was affected in the year 2000 by an outbreak of acute bacterial gastroenteritis following water contamination with *Escherichia coli* 0157:H7 and *Campylobacter jejuni* pathogen species.

# 5.2 Genetic risk factors for post-infectious irritable bowel syndrome following a waterborne outbreak of gastroenteritis

## 5.2.1 Abstract

**Background:** Acute gastroenteritis is the strongest risk factor identified to date for the development of irritable bowel syndrome (IBS), the most common and costly gastrointestinal disorder in Western society. In May 2000, over 2300 residents of a small Canadian town (Walkerton, Ontario) developed gastroenteritis (GE) from microbial contamination of the municipal water supply. An ongoing longitudinal study found that over one third of those with GE developed chronic IBS-like symptoms. We used this well-characterized cohort to study genetic susceptibility to development of post-infectious (PI)-IBS. **Methods:** We screened for 79 reported functional variants: selected candidate genes involved in serotoninergic pathways, intestinal epithelial barrier function and innate immunity. Walkerton residents who experienced GE and reported PI-IBS 2-3 years after the outbreak (229 cases) were compared to those with GE who did not develop PI-IBS (583 controls). **Results:** Four candidate variants were significantly associated with PI-IBS. Two located in *TLR9*: rs352139 (coding, P545P) ($p$=0.0135) and rs5743836 (promoter,-T1237C) ($p$=0.0324) ($r^2$<0.14); one in *CDH1*: rs16260 (promoter,-C160A) ($p$=0.0352); and one in *IL6*: rs1800795

(promoter,-G174C) ($p$=0.0489). Denser mapping was conducted in these 3 regions. We observed one novel association in *IL6* ($p$=0.0012) and 12 additional associations that can be explained by linkage disequilibrium with the 4 original variants. **Conclusion:** This is the first study to assess genetic determinants of PI-IBS. Genes encoding for epithelial cell barrier function and innate immune response to enteric bacteria are associated with the development of IBS following acute GE.

## 5.2.2 Introduction

Irritable bowel syndrome (IBS) is defined as recurrent abdominal discomfort and disturbed defecation that cannot be explained by structural or biochemical abnormalities (Thompson et al. 1999). Most people with IBS recall a gradual onset of symptoms, but some with previously normal bowel habit note an acute onset after an episode of gastroenteritis (GE) (Dunlop et al. 2003a; Dunlop et al. 2003b). After exclusion of known organic disorders, such patients are diagnosed with post-infectious irritable bowel syndrome (PI-IBS) (Spiller 2003a). A retrospective analysis reported that up to 30% of all IBS patients have an acute onset of symptoms (Chaudhary et al. 1962). The annual cost of IBS in the United States is between $1.7 billion and $10 billion in direct medical costs and $20 billion for indirect costs (Hulisz et al. 2004).

Although a link between chronic IBS-like symptoms and antecedent GE was long suspected (Chaudhary et al. 1962), the entitity of PI-IBS was not formally recognized until prospective studies identified and quantified a strong

association (McKendrick et al. 1994; Gwee et al. 1999; Spiller et al. 2000). Initial studies revealed that individuals were more likely to develop PI-IBS if they displayed a vigorous innate immune response to the infection, and inefficiently down-regulated this response after infection (Gwee et al. 2003). In addition, it was shown that patients with PI-IBS had low grade inflammation in the colon and increased intestinal permeability (Gwee et al. 1999; Spiller et al. 2000). Known risk factors for developing PI-IBS include female gender, psychological factors, severity of the acute enteric illness and microbial virulence factors (Thabane et al. 2007). However, genetic risk factors for PI-IBS have never been evaluated.

While familial aggregation and twin studies support the involvement of genetic risk factors in the pathogenesis of IBS, previous association studies have failed to identify clear and reproducible candidate genes. Limitations of these studies have included, the heterogeneity of disease, ethnicity, small sample size, poor statistical analyses, and lack of clinical and epidemiological information regarding IBS patients (Tabor et al. 2002; Saito et al. 2008a). By studying a well defined cohort of IBS patients with similar onset, etiology and symptom profile, better insight into genetic risk and/or protective factors could be achieved. Ultimately, this may assist in the development of new biomarkers, preventive strategies and treatments for this common condition.

Walkerton is a small rural town located in a prime agricultural area 180 km northwest of Toronto (Canada). In May 2000, heavy rainfall washed livestock fecal residue from nearby farms into an inadequately chlorinated shallow well. The municipal water supply was then contaminated with *Escherichia coli* 0157:H7, *Campylobacter jejuni*, and other pathogen species leading to an

outbreak of acute bacterial GE that affected over 2300 residents, with 27 recognized cases of the hemolytic uremic syndrome and 7 deaths (Marshall et al. 2006). Approximately 36% of local residents exposed to GE fulfilled Rome I diagnostic criteria for IBS when assessed two to three years after the outbreak. This unfortunate tragedy provided a unique opportunity to study the epidemiology, natural history and genetic determinants of PI-IBS in a large, well-defined study cohort with simultaneous and well characterized exposure (Marshall et al. 2006).

We undertook a study to screen candidate gene variants in this cohort. These genes were selected a priori in 3 categories: (1) reported IBS genes; (2) genes involved in intestinal epithelial barrier; and (3) genes encoding reported innate immunity risk variants. The purpose of the study was to identify polymorphisms that confer susceptibility or protection to PI-IBS and thereby to contribute to a better understanding of its pathogenesis.

### 5.2.3 Methods

**Walkerton Health Study (WHS) cohort.** The WHS was initiated in 2002 to study the epidemiological and long-term health outcomes of municipal water contamination in May 2000, and to facilitate local residents' access to specialty clinical care. Details of the WHS cohort and study methodology have been reported previously (Garg et al. 2005a; Garg et al. 2005b; Marshall et al. 2006; Garg et al. 2006). In the years 2002-2003, 4315 local residents were recruited. Participants provided blood samples for future genetic studies and were invited to

return for annual structured assessments at a local research clinic. 2069 subjects who fulfilled the following criteria were included in an PI-IBS study cohort: (1) age at least 16 years at the time of the outbreak; (2) no diagnosis of IBS, inflammatory bowel disease or celiac disease before the outbreak; and (3) permanent residency in Walkerton (as identified by postal code) at the time of the outbreak (Marshall et al. 2006). A modified version of Talley's Bowel Disease Questionaire (Talley et al. 1989) was administered to all eligible participants, and used Rome I criteria to identify those with IBS (Thompson et al. 1992). Exposure to acute GE was classified as either self-reported or confirmed, if corroborated by review of health care records, stool cultures or responses to a public health questionnaire administered during the outbreak.

**WHS cohort: Subjects in genetic study.** Of 2114 WHS participants who provided informed consent for participation in the genetic study, 1253 were eligible (113 entered the WHS after 2003, 123 were less than 16 years of age at the time of the outbreak, 163 had a prior diagnosis of IBD or IBS, and 463 were not permanent residents of Walkerton). All studies were approved by the Hamilton Health Sciences/McMaster University Faculty of Health Sciences and the McGill University Health Centre Research Ethics Boards and all participants gave written informed consent.

Eligible participants in the genetic study were divided into 4 groups for analysis: (1) those who did not experience acute GE during the outbreak and did not develop IBS (n=398); (2) those who experienced acute GE but did not develop IBS (n=581); (3) those with self-reported acute GE who subsequently

developed IBS (n=142); and (4) those with confirmed GE who subsequently developed IBS (n=86) (**Table 5.1**). Because of limited sample size, self-reported and confirmed GE was combined for analysis, although secondary analyses considered only confirmed GE.

**DNA isolation and Genotyping experiment.** DNA from peripheral whole blood was extracted using the Gentra automated system Autopure (Qiagen, Ontario, Canada) according to the manufacturers' protocols. All single nucleotide polymorphism (SNP) genotyping was performed using Sequenom iPLEX® GOLD genotyping assays (MALDI-TOF mass spectrometry) (van den Boom et al. 2007) or the TaqMan 5' exonuclease assay (Whitcombe et al. 1998) (ABI, CA, USA) (rs1800035: Assay ID C_7611968_30; rs2631367: Assay ID C_26479161_30). All primers and probes used for genotyping experiments are available in **Supplementary Table 5.1**.

DNA samples and SNPs were not included in the analysis if they did not reach high enough call rates. To be considered successful, a SNP had to have a call rate above 95% among all genotyped samples, and show a distribution of genotypes consistent with Hardy-Weinberg equilibrium (HWE) at the level $p>0.001$. For SNPs genotyped by the multiplex Sequenom assays, filtering of DNA samples based on missingness of genotype data was done on a panel-by-panel basis. For any given panel, a sample was considered failed, and all its genotypes missing, if it had a call rate below 90% among the set of successful SNPs. Samples that failed all Sequenom panels have also been considered to have failed all TaqMan assays.

All first-degree relatives of family members involved in Mendelian errors were treated as having missing genotypes at the SNPs involved in the inconsistencies, or at all SNPs if the number of inconsistencies was 4 or greater.

The significance levels that we report hereafter are calculated based upon the above filtering rules that consider all of the SNPs genotyped as part of this study, including the follow-up panels. Results that we obtained in the first exploratory phase of the study have thus been revisited once additional panels revealed more problematic samples that have been declared failed.

**Candidate genes and selection of single-nucleotide polymorphism.** With the use of public databases, including PubMed and Online Mendelian Inheritance in Man (OMIM), we selected 51 candidate genes that have been characterized and potentially associated with IBS, serotoninergic pathways (hypothesized to play a key role in IBS pathogenesis), intestinal epithelial barrier function (known to be altered in IBS and PI-IBS patients), or innate immune response, which may play key roles in the intestinal inflammatory process (see **Supplementary Table 5.2**, **Supplementary Refences**). To select the final candidates, we proceeded on the basis of a comprehensive overview of the intestinal inflammatory system biology and with the hypothesis that PI-IBS might result from persistent local low-grade inflammation. We subsequently selected 79 polymorphisms in these genes that had been previously reported in the literature as being "functional variants". Most were located in promoter regions, exons, splice-donor or splice-acceptor sites in introns or coding regions of gene, where they were expected or were reported to cause changes in function or expression level of the encoded protein.

**Fine mapping experiments: SNPs selection.** For fine mapping of candidate genes showing significant association in univariate analysis, we first performed a thorough literature research to force include in our design any SNPs that had been previously reported to be associated with different inflammatory disorders, in order to facilitate comparisons. Additionally, we included all non-synonymous SNPs reported in dbSNP (the SNP database of the National Center for Biotechnology Information) with MAF > 1%, as well as all SNPs within conserved motifs across species, SNPs disrupting putative transcription-factor, microRNA or enhancer binding sites that were all identified with the use of the University of California at Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu). We also included SNPs located in predicted regulatory regions using a computational method called ESPERR (Taylor et al. 2006) (Evolutionary and Sequence Pattern Extraction through Reduced Representations) that is also available from the UCSC genome browser under the name "7X Reg Potential" within the "Expression and Regulation" track group. To capture the remaining variations in the region of these three genes (defined as the sequence ranging from 10kb upstream of the coding sequence to 10kb downstream), we used data from the HapMap Public Release #23 and selected tagging SNPs using a pairwise tagging approach (de Bakker et al. 2005) ($r^2 \geq 0.8$).

**Statistical Analysis.** We assessed univariate association between genetic variants and PI-IBS using mixed effects logistic regression models, in which a random intercept is included to account for the correlation between individuals belonging to the same family (Xu et al. 2007). Use of logistic mixed models allows inclusion

of covariates and evaluation of multi-locus effects, in multivariate models. Analyses were done in R (http://www.r-project.org/). Allelic frequencies that we report in cases and in controls are estimates that account for relatedness (McPeek et al. 2004), as calculated by MQLS (Thornton et al. 2007).

## 5.2.4 Results

**Study cohort**

Analysis compared subjects who experienced GE but did not develop PI-IBS (581 controls; 570 successfully genotyped) to those who experienced GE and reported PI-IBS 2-3 years after the outbreak (228 cases; 220 successfully genotyped) (**Table 5.1**). Among the 790 successfully genotyped cases and controls, 533 (149 cases, 384 controls) did not have known relationship with any other individual. The remaining 257 samples (71 cases, 186 controls) belonged to 90 families.

Of 228 eligible cases, 148 were female (64.9%) compared to 311 of 581 eligible controls (53.5%). Mean age at the time of the outbreak was 44.4 years among cases versus 48.0 years among controls. A detailed description of the study cohort demographics can be found in **Table 5.2**. Features of the enteric illness included bloody stools in 33.6% of cases compared to 17.8% of controls, abdominal cramps in 98.7% compared to 84.9%, weight loss of at least 10 pounds in 31.7% compared to 14.6%, fever in 45.1% compared to 32.0%, duration of diarrhea over 7 days in 42.7% compared to 28.3%, and a number of loose stools per day greater than 10 in 31.3% compared to 20.5%.

**SNP genotyping performance**

The total number of unique SNPs assessed was 79 in the exploratory phase, and 79 in the follow-up phase. In the exploratory phase, 4 failed due to low call rates, 4 had minor allele frequency below 1% and were excluded, and none of the remaining SNPs failed HWE at the level 0.001. In the follow-up phase, these numbers of excluded SNPs are respectively 9, 7 and 1.

Out of the 2115 samples genotyped, 4 were considered to have failed all SNPs due to low call rate in all Sequenom panels. The number of samples that were considered failed for at least one of the four panels but not failed for at least another ranged from 14 to 65 (average of 39).

From all SNPs considered successfully genotyped, we identified 224 Mendelian inconsistencies, a large majority of which (216) clustered in 13 nuclear families, likely the result of mispaternity, contamination or unresolved sample switch. After removal of these nuclear families, only 8 errors remained, all involving different families.

**Exploratory phase: Candidate-SNPs screening**

Four candidates showed significant association with PI-IBS in the first stage of the study. Two located in the Toll-like receptor 9 (*TLR9*) regions are coding SNP rs352139 (P545P) ($p$=0.0116; OR=1.338; CI: 1.067-1.678; frequency of allele A: 48% in cases, 41% in controls) and promoter SNP rs5743836 (-T1237C) ($p$=0.0291; OR=0.701; CI: 0.509-0.964; frequency of allele T: 87% in cases, 82% in controls) (**Supplementary Table 5.2**). These 2 variants are 2.4kb apart and are in low linkage disequilibrium with each other (r2<0.14). The third

associated variant, rs16260 (-C160A) ($p$=0.0642; OR=1.258; CI: 0.986-1.605; frequency of A allele: 31% in cases, 26% in controls), is located in *CDH1* promoter region. The variant rs16260 was initially significantly associated upon the exploratory association analyses, suggesting CDH1 as a PI-IBS candidate gene, but we report here the result of the final analysis that combines the fine mapping experiment results (Methods). The fourth candidate, rs1800795 (-G174C) ($p$=0.0309; OR=1.302; CI: 1.024-1.655; frequency of C allele: 44% in cases, 39% in controls), is located in the promoter region of *IL6*. These 4 associations were evaluated further in subsequent fine-mapping experiments.

**Fine Mapping of *TLR9*, *IL6* and *CDH1* regions**

Fine mapping experiments used the tagging strategy described above (Methods). The screened *TLR9* region (44.9kb) included a cluster of other coding genes comprising *ALAS1*, *TWF2*, *PPM1M* and *WDR82* (**Figure 5.1A**). Among 11 SNPs successfully genotyped in this region, 5 are significantly associated with PI-IBS status. However, all of these can solely be explained by their level of LD with the 2 associated SNPs from the exploratory phase (rs352139 and rs5743836). Given that these 2 SNPs are in weak LD with each other, the additional associations observed in the region are divided in two clusters of associations (**Figure 5.1A**). The first includes SNPs in LD with rs352139 (rs352163, rs352140, rs1060330), where the minor allele of all 4 SNPs are associated with an increase risk in developing PI-IBS (rs352163: $p$=0.00575) (**Supplementary Table 5.3**). The second cluster of association includes SNPs in LD with rs5743836 (rs11717574,

rs4082828), where the major allele of all 3 SNPs are associated with a decrease risk in developing PI-IBS (rs4082828: $p$=0.0170) **(Supplementary Table 5.3)**.

The studied *CDH1* region (147kb) also comprised part of the coding sequence of *TMCO7* (promoter to intron 1) **(Figure 5.2A)**. Among the 27 SNPs successfully genotyped in this region, 4 additional SNPs (one located in *CDH1* promoter (rs7186693) and 3 in intron 2 (rs2010724, rs10431923, rs7186053)) are significantly associated with PI-IBS **(Supplementary Table 5.3)**. As for TLR9, these associations can solely be explained by their level of LD with the *CDH1* promoter SNP rs16260 associated in the exploratory phase, where the minor allele of all 5 SNPs confers increased risk to developing PI-IBS (rs7186053: $p$=0.0218) **(Supplementary Table 5.3)**.

A window of 29kb was screened in the *IL6* region, which only included the coding sequence of *IL6* **(Figure 5.3A)**. Among the 24 successfully genotyped SNPs, the minor allele of 4 SNPs (rs4552807, rs1554606, rs2069845, rs2069861) are additionally associated with increase risk of developing PI-IBS **(Supplementary Table 5.3)**. The observed associations with three of these SNPs (rs4552807, rs1554606, rs2069845) (rs4552807: $p$=0.0134) can be explained by their level of LD with the promoter variant rs1800795 associated in the exploratory phase of the study. SNP rs2069861, located 34bp downstream of *IL6* 3' UTR, is independently associated with an increase risk of developing PI-IBS ($p$= 0.00788; OR= 1.60814; CI: 1.13274- 2.28305; frequency of T allele: 14% in cases, 9% in controls). This latter SNP tagged only for itself in our design as well as over 1Mb when compared with SNPs from HapMap data from Public Release #23.

Overall, these association results are robust to the possible misclassification of cases designed as "probable GE". Indeed, when association analyses in these 3 candidate gene regions are limited to confirmed GE cases (172 successfully genotyped controls vs. 83 successfully genotyped cases), variants in both the *CDH1* **(Figure 5.2B)** and the *IL6* **(Figure 5.3B)** regions are even more significantly associated than in the analysis including both the self-reported (570 controls vs. 220 cases) and confirmed cases **(Supplementary Table 5.4)**. The *TLR9* variants did not show greater association level in the confirmed sample set **(Figure 5.1B) (Supplementary Table 5.4)**.

In the case of *CDH1*, 11 additional SNPs are associated with confirmed PI-IBS **(Figure 5.2B, Supplementary Table 5.4)**, 3 of which are located in the intron 2 region, bringing to four the number of associated SNPs in *CDH1* intron 2 (rs4783681, rs1125557, rs12597188, rs2010724). The minor allele of 7 associated SNPs (rs7186693, rs4783681, rs1125557, rs12597188, rs2010724, rs10431923 and rs7186053) located between the promoter and intron 3 and all in high LD with promoter variant rs16260, confers increase risk to developing PI-IBS in the subset of confirmed GE cases **(Figure 5.2B, Supplementary Table 5.4)**. The strongest association signal is observed with rs2010724, located in *CDH1* intron 2 ($p$= 0.00008; OR= 2.142; CI: 1.465- 3.131; frequency of C allele: 44.6% in cases, 25.0% in controls), a level that withstand a Bonferroni correction for the total number of SNPs tested ($p_{corr}$=0.011, accounting for 133 SNPs). A second cluster of association with a decrease risk of developing PI-IBS is observed with the major allele of 8 SNPs (rs4783573, rs7188750, rs2276330, rs8061932, rs1801026, rs13689, rs7197744 and rs6499203) that are all in LD and

are located between *CDH1* intron 3 and 3'UTR region (**Figure 5.2B**, **Supplementary Table 5.4**). The best association signal is observed with rs6499203, located 16kb downstream from *CDH1* ($p$=0.00348; OR=0.494; CI: 0.308-0.793; frequency of T allele: 82.5% in cases, 69.1% in controls).

The SNP rs2069861 ($p$=0.00026; OR= 2.839; CI: 1.622-4.970; frequency of T allele: 20% in cases, 7.2% in controls) located in *IL6* (**Figure 5.3B**, **supplementary Table 5.4**) distinguishes itself from the other SNPs in this region by showing the greatest increase in risk for developing PI-IBS in the confirmed GE sample set. As mentioned previously, this SNP is independent from other associations observed in the region and tagged only itself in our design.

### *TLR9*, *CDH1* and *IL6* are independent PI-IBS risk factors

A previous epidemiological study of the Walkerton outbreak looking at 2069 eligible individuals affected with GE identified younger age, female gender, bloody stools, abdominal cramps, weight loss, and prolonged diarrhea as independent risk factors for developing PI-IBS (Marshall et al. 2006). When this multiple logistic regression was repeated for the 1253 individuals from this cohort who provided DNA for genetic analysis, the same independent risk factors were observed (**Table 5.3**). We subsequently evaluated and concluded that all of the identified genetic risk factors in the *TLR9*, *IL6* and *CDH1* regions are independent risk factors for developing PI-IBS when controlling for previously identified clinical risk factors (**Table 5.4**).

## 5.2.5 Figures

**Figure 5.1: Compilation of association results of *TLR9* region**

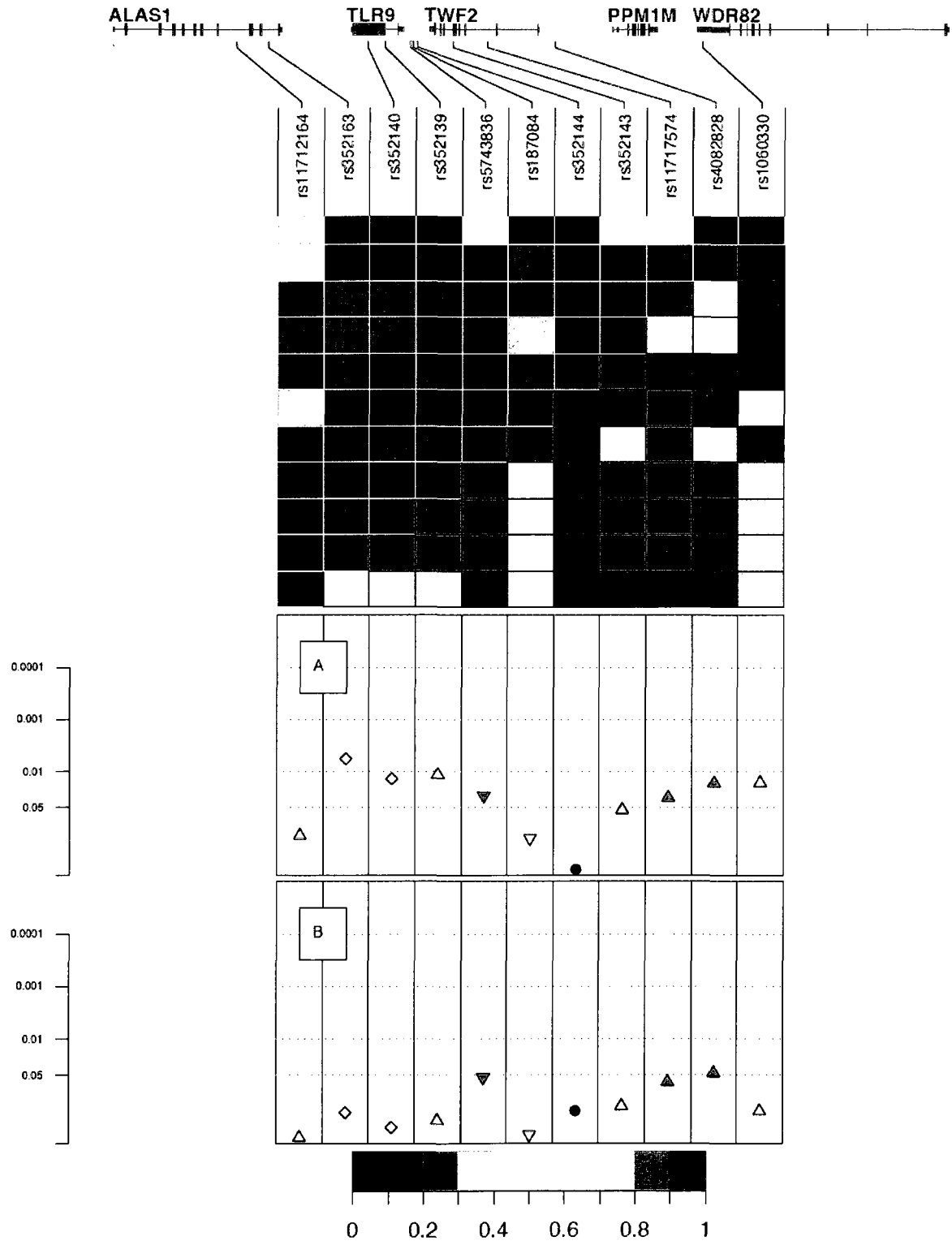


**Figure 5.1:** In the upper panel are the SNPs with their positions in the genes and the LD structure between them. SNP names in red were genotyped in the second phase of the study, subsequently to the sequencing experiment. The upper left portion is *D'* and the lower right portion is $r^2$. In the lower panel are the results from association analysis of all PI-IBS patients (**panel A**) and confirmed PI-IBS patients (**panel B**). *P*-values for individual alleles are reported in a logarithmic scale on the Y-axis, with symbols representing the associated allele (Δ=A, o=C, ▼=T, ◊=G). The color spectrum is used to represent the strength of LD and the frequency of the associated alleles.

**Figure 5.2: Compilation of association results of *CDH1* region**

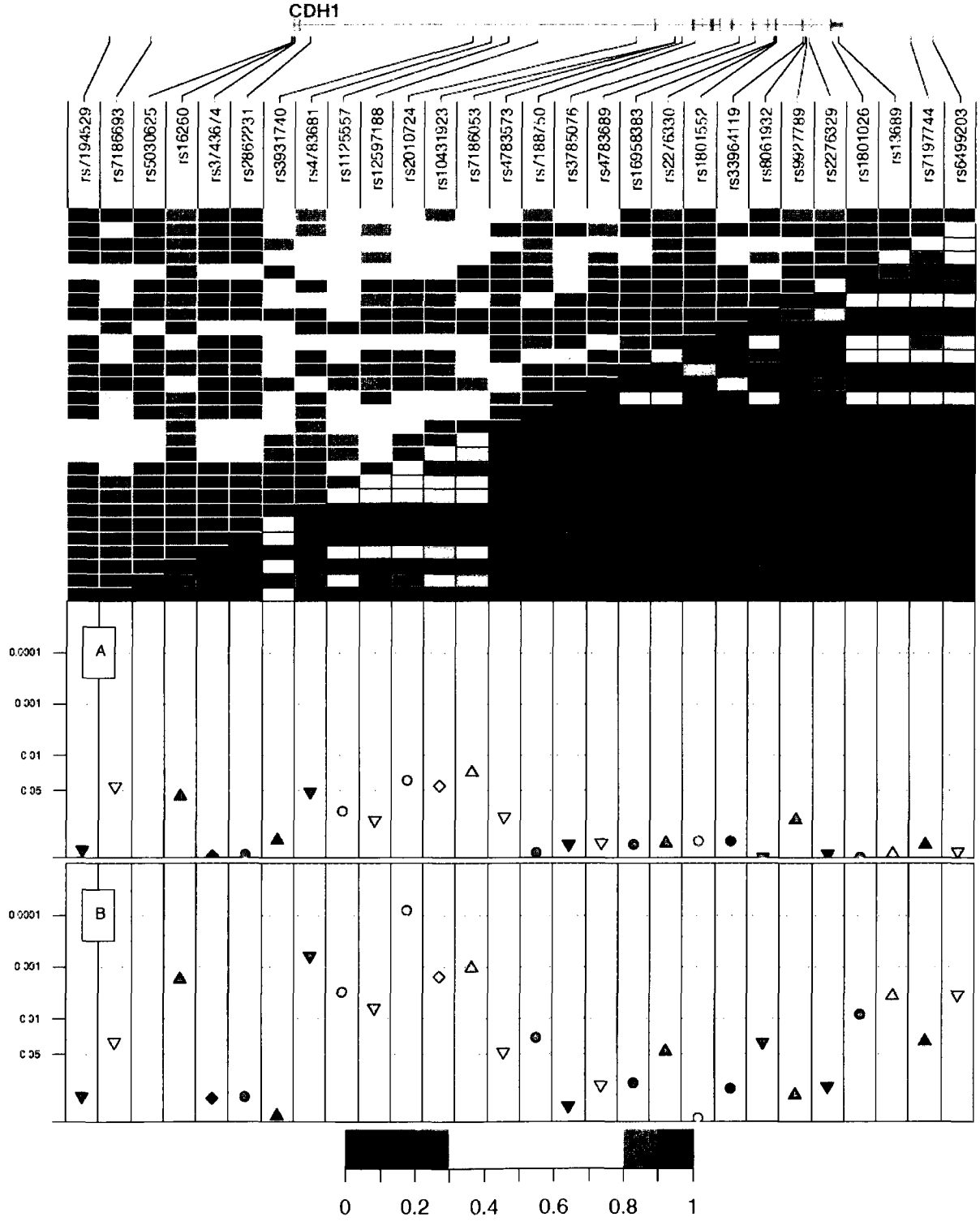


**Figure 5.2:** In the upper panel are the SNPs with their positions in the genes and the LD structure between them. SNP names in red were genotyped in the second phase of the study, subsequently to the sequencing experiment. The upper left portion is *D'* and the lower right portion is $r^2$. In the lower panel are the results from association analysis of all PI-IBS patients (**panel A**) and confirmed PI-IBS patients (**panel B**). *P*-values for individual alleles are reported in a logarithmic scale on the Y-axis, with symbols representing the associated allele (Δ=A, o=C, ▼=T, ◊=G). The color spectrum is used to represent the strength of LD and the frequency of the associated alleles.

**Figure 5.3: Compilation of association results of *IL6* region**

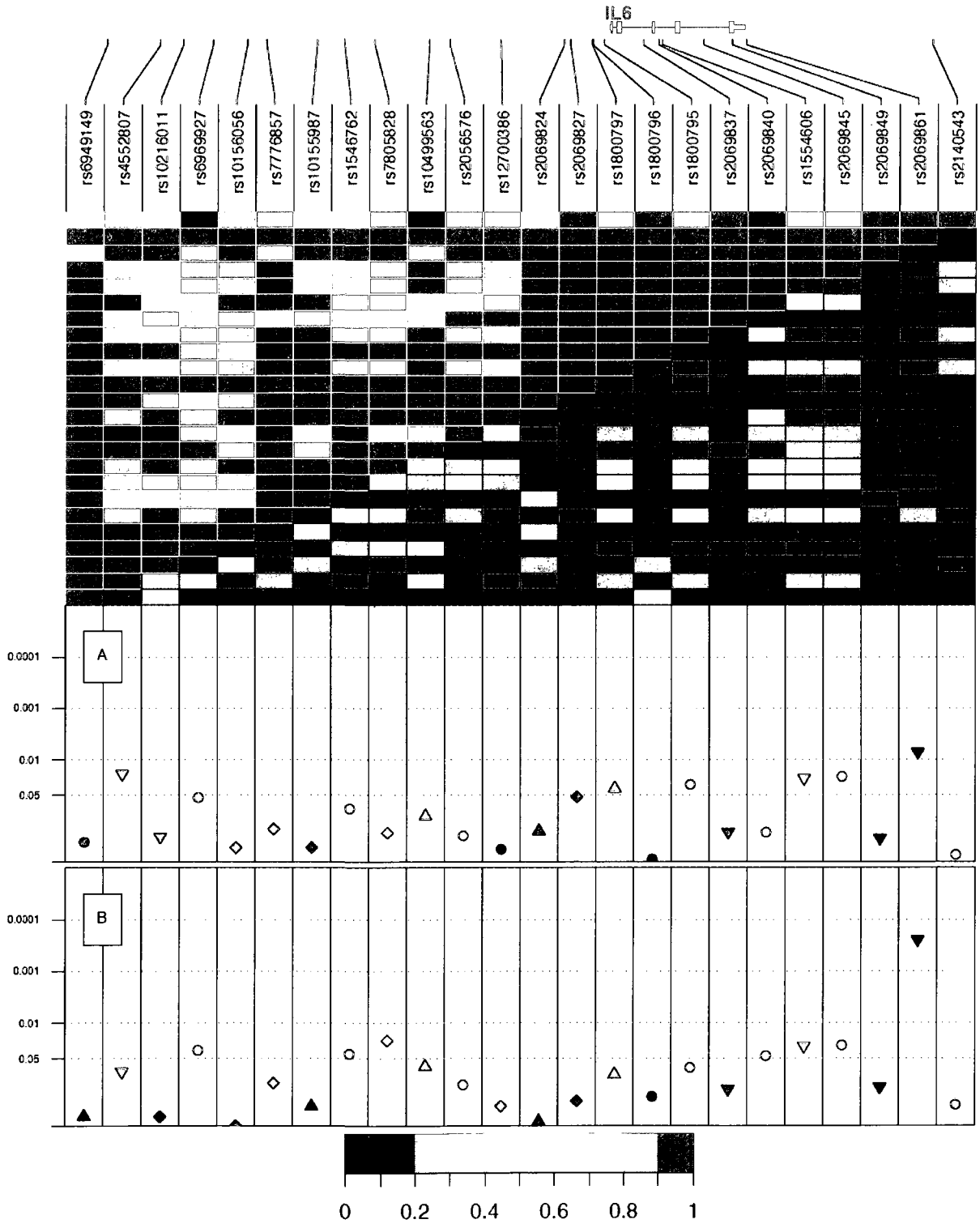


**Figure 5.3:** In the upper panel are the SNPs with their positions in the genes and the LD structure between them. SNP names in red were genotyped in the second phase of the study, subsequently to the sequencing experiment. The upper left portion is *D'* and the lower right portion is $r^2$. In the lower panel are the results from association analysis of all PI-IBS patients (**panel A**) and confirmed PI-IBS patients (**panel B**). *P*-values for individual alleles are reported in a logarithmic scale on the Y-axis, with symbols representing the associated allele (Δ=A, o=C, ▼=T, ◊=G). The color spectrum is used to represent the strength of LD and the frequency of the associated alleles.

### 5.2.6 Tables

**Table 5.1: Walkerton PI-IBS cohort description**

| | | Exposure | | | |
|---|---|---|---|---|---|
| | Category | Control | Suspected GE | Confirmed GE | Total GE |
| **PI-IBS** | No | 398 | 407 | 174 | 979 |
| | Yes | 46 | 142 | 86 | 274 |
| | Total | 444 | 549 | 260 | 1253 |

**Table 5.2: Demographic characteristics of the Walkerton PI-IBS cohort**

| | PI-IBS | |
|---|---|---|
| | **Yes (cases)** | **No (controls)** |
| **Number of individuals** | 228 | 581 |
| **Demographic** | | |
| Gender (%) | | |
| Male | 80 (35.09) | 270 (46.5) |
| Female | 148 (64.91) | 311 (53.5) |
| Age at the time of the outbreak mean±s.d. (range) | 44.4±14.0 (16.4-73.8) | 48.0±16.6 (16.0-90.9) |
| **Features of enteric illness at time of the outbreak** | 74 (33.6) | 91 (17.8) |
| Bloody stools (%) | 8 | 69 |
| Not reported | 224 (98.7) | 485 (84.9) |
| Abdominal cramps (%) | 1 | 10 |
| Not reported | 72 (31.7) | 83 (14.6) |
| Weight loss (> 10 pounds) (%) | 1 | 11 |
| Not reported | 102 (45.1) | 182 (32.0) |
| Fever | 2 | 12 |
| Not reported | | |
| Prolonged diarrhea (days) (%) | | |
| 0-1 | 9 (4.1) | 50 (9.7) |
| 2-3 | 40 (18.2) | 151 (29.4) |
| 4-5 | 42 (19.1) | 90 (17.5) |
| 6-7 | 35 (15.9) | 77 (15.0) |
| >7 | 94 (42.7) | 145 (28.3) |
| Not reported | 8 | 68 |
| Maximum no. of stools/day | | |
| 1-3 loose stools/day | 26 (12.3) | 95 (19.7) |
| 4-6 loose stools/day | 61 (28.9) | 196 (40.6) |
| 7-10 loose stools/day | 58 (27.5) | 93 (19.3) |
| > 10 loose stools/day | 66 (31.3) | 99 (20.5) |
| Not reported | 17 | 98 |

**Table 5.3: Multiple logistic regression analysis of clinical predictors of IBS after acute gastroenteritis using the 1253 patients who provided DNA for the genetic analysis**

| Variables | $\beta$ coefficient (standard error) | *P* value | Odds ratio (95% confidence interval) |
|---|---|---|---|
| Age, *y* | -0.012 (0.006) | 0.0340 | 0.988 (0.978-0.999) |
| Female gender | 0.461 (0.181) | 0.0109 | 1.586 (1.112-2.261) |
| ***Features of acute enteric illness*** | | | |
| Duration of diarrhea (days) | | | |
| 0-1 | Reference | | |
| 2-3 | -0.006 (0.420) | 0.988 | 0.994 (0.436-2.265) |
| 4-5 | 0.351 (0.429) | 0.413 | 1.420 (0.613-3.290) |
| 6-7 | 0.298 (0.439) | 0.496 | 1.348 (0.570-3.185) |
| >7 | 0.638 (0.409) | 0.118 | 1.893 (0.850-4.216) |
| Bloody stools | 0.576 (0.204) | 0.00465 | 1.779 (1.194-2.651) |
| Abdominal cramps | 2.037 (0.606) | 0.00078 | 7.671 (2.337-25.179) |
| Weight loss (> 10 pounds) | 0.740 (0.213) | 0.000518 | 2.095 (1.380-3.181) |
| Constant (Intercept) | -3.175 | | |

**Table 5.4: Multiple logistic regression analysis of clinical and genetic predictors of IBS after acute gastroenteritis**

| Variables | β coefficient (standard error) | *P* value | Odds ratio (95% confidence interval) |
|---|---|---|---|
| rs5743836 (*TLR9*) | 0.420 (0.179) | 0.0191 | 1.521 (1.071-2.161) |
| rs2069861 (*IL6*) | 0.402 (0.194) | 0.0385 | 1.494 (1.022-2.186) |
| rs16260 (*CDH1*) | 0.329 (0.137) | 0.0160 | 1.390 (1.063-1.816) |
| Age, *y* | -0.013 (0.006) | 0.0219 | 0.987 (0.976-0.998) |
| Female gender | 0.448 (0.185) | 0.0153 | 1.566 (1.090-2.249) |
| ***Features of acute enteric illness*** | | | |
| Duration of diarrhea (days) | | | |
| 0-1 | Reference | | |
| 2-3 | -0.043 (0.424) | 0.919 | 0.958 (0.417-2.200) |
| 4-5 | 0.358 (0.432) | 0.408 | 1.430 (0.613-3.335) |
| 6-7 | 0.196 (0.444) | 0.659 | 1.217 (0.510-2.906) |
| >7 | 0.598 (0.413) | 0.148 | 1.819 (0.809-4.087) |
| Bloody stools | 0.611 (0.208) | 0.003287 | 1.843 (1.226-2.770) |
| Abdominal cramps | 2.035 (0.607) | 0.000803 | 7.650 (2.328-25.139) |
| Weight loss (> 10 pounds) | 0.731 (0.218) | 0.000798 | 2.077 (1.355-3.184) |
| Constant (Intercept) | -4.066 | | |

### 5.2.7 Discussion

Using an exploratory panel of 71 successfully genotyped putative functional variants selected from the literature, we identified 3 promoter (i.e. rs5743836, rs16260, and rs1800795) and 1 coding (i.e. rs352139) variants associated with susceptibility to PI-IBS. These associations led us subsequently to focus fine mapping experiments in 3 gene regions of interest, namely *TLR9*, *CDH1* and *IL6*. This study establishes the first pillars of the genetic architecture of PI-IBS by demonstrating that variants in these 3 genes all contribute to PI-IBS susceptibility, and are all independent risk factors for developing PI-IBS when controlling for previously reported PI-IBS clinical risk factors in the Walkerton population (Marshall et al. 2006). The remarkable observation of clusters of SNPs in LD showing similar level of associations in all three genes, as well as the enhanced associations of these markers in the subset of confirmed PI-IBS individuals, further support and strengthen our genetic findings.

The coding variant P545P (rs352139) and promoter variant -T1237C (rs5743836) of *TLR9* were both associated with increase risk of developing PI-IBS. These two variants were originally selected in our exploratory design since they had been reported to distinguish the four common haplotypes of *TLR9*, accounting for more than 75% of the chromosome (Lazarus et al. 2003). Additional associations observed in the *TLR9* region, obtained by performing denser mapping, could solely be explained by their level of LD with these 2 SNPs. The *CDH1* promoter variant rs16260 (-C160A), also associated with PI-IBS susceptibility in our study, was included in the exploratory panel because it

had previously been reported to influence *CDH1* transcriptional activity. More specifically, the "A" allele associated with PI-IBS susceptibility has been reported to result in a 68% decrease in transcription efficiency compared to the "C" allele (Li et al. 2000). An additional 4 SNPs, one located in *CDH1* promoter (rs7186693) and 3 in intron 2 (rs2010724, rs10431923, rs7186053), were also found to be associated with PI-IBS susceptibility. Yet, like for *TLR9*, all of these associations could be explain by the level of LD between these 4 SNPs and promoter SNP rs16260. Interestingly, Stemmler *et al.* has reported putative cis-regulatory elements located in *CDH1* intron 2, which is one of the associated regions with PI-IBS (Stemmler et al. 2005). When deleting this entire intron in embryonic stem cells, they demonstrated that intron 2 sequences are required for both the initiation and maintenance of *CDH1* transcriptional activity (Stemmler et al. 2005). More recently, a meta-analysis of two genome-wide association studies on colorectal cancer (CRC), comprising 13,315 individuals, has identified rs9929218, located in *CDH1* intron 1 and in high LD with our signal in intron 2 ($r^2$>0.957), as a novel risk factor for CRC (COGENT Study 2008). This latter result further emphasizes the importance of *CDH1* intron 2 region, which, together with the -C160A promoter variant, might contribute to *CDH1* regulation in colonic epithelial cells. Finally, the *IL6* the promoter variant rs1800795 (-G174C) was included in the exploratory phase because the "G" allele has been reported to achieve an overall higher rate of transcription (Fishman et al. 1998). Yet, there are numerous inconsistencies within and between *in vitro* studies concerning the effect that this promoter variant has on gene expression (Heesen et al. 2002; Qi et al. 2006). This may suggest a more complex regulatory haplotype

extending further upstream or downstream from the previously analysed promoter region. Our study seems to corroborate this hypothesis as we uncovered 2 independent association signals in this study. The first one included 4 associated SNPs (rs4552807, rs1554606, rs2069845, rs2069861) all in LD with promoter variant rs1800795, and the second association signal was strictly isolated to variant rs2069861, located 34bp from *IL6* 3'UTR.

The three candidates described above (namely *TLR9*, *IL6*, and *CDH1*) represent an interplay of genetic factors with key roles the regulation of the intestinal mucosal immune response. Given the nature of the Walkerton water contamination, *TLR9* is an interesting candidate that might have played a critical role in recognizing the initial microbial threat and initiating the intestinal innate immune response. Indeed, *TLR9* is the only human pattern recognition receptor for bacteria detecting unmethylated CpG dinucleotide (Wagner et al. 1999). The recognition of such sequences activates signaling cascade via transcription factors such as NFκB and AP-1 and stimulate the proliferation of B cells and secretion of proinflamatory cytokines (i.e. IL-6, IL-12, and TNFα) that are required to eliminate invading pathogens (Krieg et al. 2002). Colonic epithelial cells are constantly exposed to high levels of bacterial DNA in the gut lumen and must recognize and respond appropriately to pathogens while maintaining a tolerance to non-pathogenic commensal bacterial strains. In that context, *TLR9* surface localization and expression have been reported to be up-regulated on intestinal epithelial cells in response to pathogenic bacteria, suggesting that the epithelial inflammatory response to pathogenic DNA is mediated, at least in part, by increased *TLR9* expression (Ewaschuk et al. 2007). Consequently, we hypothesize

that a defect in properly recognizing the initial threat could contribute to the poorly controlled innate response to the infection documented in those who develop PI-IBS (Gwee et al. 2003). Additionally, enteric pathogens, such as *Escherichia coli* and *Campylobacter*, both present in the outbreak, have devised several ways to increase mucosal permeability, mainly through disruption epithelial tight junctions (TJ) (Wu et al. 2007; Frank et al. 2007; Sasaki et al. 2007) like *CDH1*, another genetic risk factor for PI-IBS. *CDH1* is the principal mediator of cell adhesion between intestinal epithelial cells and is found in adherents junctions (AJ). The AJ and the TJs form the apical junctional complex (AJC) that is responsible for the normal barrier function in the intestine. Overall, the *CDH1* genetic association results are consistent with observations that PI-IBS patients have an increased small intestinal permeability (Spiller et al. 2000; Marshall et al. 2004; Dunlop et al. 2006). In a substudy of the WHS cohort, intestinal permeability was increased in 35% of subjects with IBS compared to just 13% in non-IBS controls (Marshall et al. 2004). Underlying defects in the AJC have long been hypothesized to cause increased mucosal permeability, and the increased intestinal permeability in PI-IBS patients may reflect TJ disruption by enteric pathogens during the acute phase of infection (Barbara 2006a).

Yet, while immune activation observed in PI-IBS patients could be caused by increased intestinal permeability, it is also plausible that inflammation and immune activation disrupt the mucosal barrier, allow luminal contents to pass into the submucosa and create a vicious inflammatory circle of immune triggers (Barbara 2006a). For example, peripheral immune cells are known to significantly interact with intestinal epithelial cells (Bisping 2001), and cytokines

released from PBMCs may modulate intestinal epithelial barrier function (Cario et al. 1999). Amongst these secreted cytokines, IL-6, a pleiotropic cytokine and main mediator of acute inflammatory responses, has been reported to reflect both the severity of intestinal inflammation (Nakamura et al. 1992) and disease activity (Holtkamp et al. 1995). Previous report showing that PBMCs isolated from IBS patients secrete a higher baseline and LPS-induced IL-6 levels compared to healthy controls (Liebregts et al. 2007) further supports the associations we observed between PI-IBS and *IL6* variants.

Because the Walkerton PI-IBS population cohort is unique in its size, in the number of DNA samples available and in the collection and follow-up of clinical data, no similar PI-IBS cohorts exist to replicate our findings. Another limitation of this study is the way we selected our candidate SNPs. We can only exclude the involvement of the reported functional and tagging SNPs that we have evaluated in PI-IBS, as well as SNPs in high LD and tagged by them. However, we cannot exclude the possible involvement of these candidate genes evaluated in the exploratory phase unless thorough fine mapping experiments is performed in all of these regions. An additional limitation of the study design is consequent to constraints on local resources during the initial Walkerton outbreak, where only a minority of local residents was encouraged to submit stool specimens for culture, explaining in part the small number of confirmed PI-IBS patients in our study. Hence, it remains possible that we have some unconfirmed cases that dilute and influence the association signals observed when combining the self-reported and confirmed patient sample sets in our analyses.

Results from the multiple logistic regression analysis suggest that independent clinical predictors and genetic risk factors contribute to the PI-IBS pathophysiology. PI-IBS is likely to be a complex trait where the variability in clinical manifestation and presentation of symptoms could be partly explained by the heterogeneity of the underlying genetic and environmental risk factors. In the case of the Walkerton PI-IBS population cohort, we had the advantage of working with a more homogeneous cohort derived from a single community, where the environmental trigger was known and where we had access to a wide-range of clinical and epidemiological data. The identification of genetic factors involved in PI-IBS is important as it provides new insight into the pathophysiology of this disorder. The identified genetic risk factors encode epithelial barrier function and innate immune response to enteric bacteria. These are in keeping with pathophysiologic changes observed in patients with PI-IBS, which include increased intestinal permeability, low grade inflammation, and immune activation in the colonic mucosa. Altogether, these observations indicate that genetic risk factors are involved in the development of IBS following acute GE.

### 5.2.8 Accession numbers

GenBank: TLR9, AF259262; CDH1, NM_004360; IL6, M18403.OMIM: *TLR9*, 605474; *CDH1*, 192090; *IL6*, 147620.

### 5.2.9 Aknowledgement

We thank the study participants and their families for their collaboration; The Génome Québec & Montreal Heart Institute Pharmacogenomics Centre for their help in isolating DNA; and A. Montpetit and the genotyping platforms of the McGill University and Genome Québec Innovation Centre for their technical assistance. This research was supported by grants to JKM from the Crohn's & Colitis Foundation of Canada and the Ontarian Ministry of Health and Long-Term Care. A.-C.V. is supported by a PhD training award from the Fonds de la recherche en santé du Québec. None of the funding agencies had any role in the design and conduct of the study; in the collection, analysis and interpretation of the data; or in the preparation, review or approval of the manuscript.

## 5.2.10 Supplementary Tables

**Supplementary Table 5.1: List of primers and probes used for genotyping experiments**

| SNP_ID | PCR Primer Forward Sequence | PCR Primer Reverse Sequence | Probe Sequence | Probe Orientation | Sequenom Panels |
|---|---|---|---|---|---|
| rs2188962 | ACGTTGGATGTTCGTGGCTGCCATAAACTG | ACGTTGGATGATGTGATACGGCGTGTAGTC | TCTTGCTCTCTGACCC | F | Panel I_Exploratory |
| rs9292777 | ACGTTGGATGGCCCTGACTCACTCTTTGAC | ACGTTGGATGAGCTCCTTCCTTATGACAAG | CCCGTAGGTTCCCCAA | F | Panel I_Exploratory |
| rs6958571 | ACGTTGGATGTGAGGTTGGGTGAGTAGAAG | ACGTTGGATGGTCCTTCTGGTGTACTGATG | CAACCTGCTGTGTGTG | R | Panel I_Exploratory |
| rs2043211 | ACGTTGGATGGCTTTTCTTACCTGGGAATG | ACGTTGGATGGAAGATGATGAGACAGAGGC | gTCAGGAACAGCACGGA | F | Panel I_Exploratory |
| rs1545620 | ACGTTGGATGTAGAGCTTCCGCTGCCAGTA | ACGTTGGATGCACTTCCTGCAGATGAAGCG | tCCAGTAGCCCCTCCATG | F | Panel I_Exploratory |
| rs2066847 | ACGTTGGATGCTTACCAGACTTCCAGGATG | ACGTTGGATGCCAGGTTGTCCAATAACTGC | TGTCATTCCTTTCAAGGG | R | Panel I_Exploratory |
| rs1050152 | ACGTTGGATGCTACATCGTCATGGGTAGTC | ACGTTGGATGCATCTGCTCTAAGGTTTCTG | GACTGTCCTGATTGGAATC | F | Panel I_Exploratory |
| rs3814055 | ACGTTGGATGGAGACCACGATTGAGCAAAC | ACGTTGGATGAGGAAGTCCAAATGTTCACC | AGCAAACAGGTAGAAAAGA | R | Panel I_Exploratory |
| rs1143627 | ACGTTGGATGCCTCGAAGAGGTTTGGTATC | ACGTTGGATGATTTCTCAGCCTCCTACTTC | cTCTCCCTCGCTGTTTTTAT | R | Panel I_Exploratory |
| rs7078219 | ACGTTGGATGGTCTAGTGGACTACCTCTTC | ACGTTGGATGTCTGAATCCCCATTCCTACC | aAAACAATGTGGTCTTCAAC | F | Panel I_Exploratory |
| rs1143634 | ACGTTGGATGGTGCTCCACATTTCAGAACC | ACGTTGGATGGCAGTTCAGTGATCGTACAG | CATTTCAGAACCTATCTTCTT | F | Panel I_Exploratory |
| rs11209026 | ACGTTGGATGGGTGATACATATACATGTAG | ACGTTGGATGGGGAATGATCGTCTTTGCTG | aCTGCAAAAACCTACCCAGTT | R | Panel I_Exploratory |
| rs13361189 | ACGTTGGATGATGGACAGTCAGTACCCTGC | ACGTTGGATGTCTAAACTGTACCCGCTGAG | cAGCAGAGTGTGCTTGAAAAT | F | Panel I_Exploratory |
| rs28362491 | ACGTTGGATGTAGGGAAGCCCCCAGGAAG | ACGTTGGATGATGGACCGCATGACTCTATC | CCCAGGAAGCGCCTGCCGGGCC | R | Panel I_Exploratory |
| rs224135 | ACGTTGGATGGTTCTGATTGGCAATAGAGG | ACGTTGGATGTTCCACAGTGAAAGTGTGGC | TTAGGGAAAAGTTTACCTATGT | F | Panel I_Exploratory |
| rs1248696 | ACGTTGGATGTGAATGGAGAGGTTCTCCAC | ACGTTGGATGATTATTCCCCTTCCACAGGC | gCTCCACCTTCTCATTCACTTGC | F | Panel I_Exploratory |
| rs4958847 | ACGTTGGATGCTATTTTAGGGAGTAGTGC | ACGTTGGATGGGGTGAGGTCATCTTAACAG | aAGGGAGTAGTGCATTATTTAGC | R | Panel I_Exploratory |
| rs4986790 | ACGTTGGATGTAGAGGGCCTGTGCAATTTG | ACGTTGGATGCACACTCACCAGGGAAAATG | CATACTTAGACTACTACCTCGATG | F | Panel I_Exploratory |
| rs1800872 | ACGTTGGATGTCCTCAAAGTTCCCAAGCAG | ACGTTGGATGAAAGGAGCCTGGAACACATC | gTTCCAGAGACTGGCTTCCTACAG | R | Panel I_Exploratory |
| rs10045431 | ACGTTGGATGGGCCAAGGCATCAATTTGAG | ACGTTGGATGTCTTGTTCCTCTTCTGGCTG | gAGGCATCAATTTGAGAGTTTAAT | R | Panel I_Exploratory |
| rs2076756 | ACGTTGGATGCTTGTCCTCTCAGGGTTTTG | ACGTTGGATGGCACCCACTACCAATGGATT | tGATCAGGTACATTTTATCTTAAGG | F | Panel I_Exploratory |
| rs6887695 | ACGTTGGATGTGTGCTTCTCGGTAAGTCAG | ACGTTGGATGGGTCACAAGCGTAGTAAATG | GAGAGAAGCAGTGTAGTGTAGTGGT | F | Panel I_Exploratory |
| rs224136 | ACGTTGGATGCATTCATCCAAGTGCCCAAG | ACGTTGGATGATAGGGCTGTTCACAACTGG | AGTGCCCAAGTGAAGAACTTGAGTGC | R | Panel I_Exploratory |

| | | | | | |
|---|---|---|---|---|---|
| rs3789243 | ACGTTGGATGTCTCTGACTGCTTCAGTTCC | ACGTTGGATGACCCAGACAATAAGCCCAAG | TGACTGCTTCAGTTCCAACAACGACGC | F | Panel I_Exploratory |
| rs4613763 | ACGTTGGATGTCTTCTGCATAGGGATCAGG | ACGTTGGATGTTCTGTGTTGGGTCTCTCTG | tTCTGATATAAGACAGGTTTATTCCCA | F | Panel I_Exploratory |
| rs7645243 | ACGTTGGATGTGGTGTCTCGGATGGATAAG | ACGTTGGATGAGCCCACACAGCTAAGAAAG | tATAAGGTGGAACGCCTGAAACAGGTG | R | Panel I_Exploratory |
| rs2847297 | ACGTTGGATGATATACGGGAAAACCTGGGC | ACGTTGGATGGAATGACAAGACTACTCTCG | cTATACGGGAAAACCTGGGCAGGTCCCA | R | Panel I_Exploratory |
| rs2241880 | ACGTTGGATGCTGTCTCTTCCTTCCCAGTC | ACGTTGGATGCTTACGAAGACACACAAGGC | CCTTCCCAGTCCCCCAGGACAATGTGGAT | R | Panel I_Exploratory |
| rs5744168 | ACGTTGGATGTCCTGGAAAAATTACAGACC | ACGTTGGATGGCCACTCAAGAAGATATCGG | gCCTGGAAAAATTACAGACCTTGGATCTC | F | Panel I_Exploratory |
| rs1045642 | ACGTTGGATGTATGTTGGCCTCCTTTGCTG | ACGTTGGATGCCTGTTTGACTGCAGCATTG | cCCTTTGCTGCCCTCAC | R | Panel I_Exploratory |
| rs2066844 | ACGTTGGATGAGTGCCAGACATCTGAGAAG | ACGTTGGATGGGATGGAGTGGAAGTGCTTG | ccTGAGAAGGCCCTGCTC | F | Panel I_Exploratory |
| rs2066845 | ACGTTGGATGCTGACACTGTCTGTTGACTC | ACGTTGGATGGTGATCACCCAAGGCTTCAG | ccGCCTTTTCAGATTCTGG | F | Panel I_Exploratory |
| rs1800630_rs1799724#1 | ACGTTGGATGATGGGTAGGAGAATGTCCAG | ACGTTGGATGTCCCCTGTATTCCATACCTG | ttggTCGAGTATGGGGACCCCC | F | Panel I_Exploratory |
| rs1800630_rs1799724#2 | ACGTTGGATGATGGGTAGGAGAATGTCCAG | ACGTTGGATGTCCCCTGTATTCCATACCTG | ccccaggACATGGCCCTGTCTTC | R | Panel I_Exploratory |
| rs1799964 | ACGTTGGATGGGGAAGCAAAGGAGAAGCTG | ACGTTGGATGTGTAACCCATTCCTCAGAGC | tgAAGGAGAAGCTGAGAAGA | F | Panel I_Exploratory |
| rs4495224 | ACGTTGGATGTCTGCCTGTAAGGTCCAATC | ACGTTGGATGAGGAGGTGAAGTCCTTGAAC | cccagACCCACAGAGTTTAAATTGG | F | Panel I_Exploratory |
| rs2542151 | ACGTTGGATGGCAAGAAGGTGTGAAGTTAG | ACGTTGGATGTCCCAAACTCTAGATGCCAC | caggGAAAATGAGAGTCTCAGGAAG | R | Panel I_Exploratory |
| rs1800629 | ACGTTGGATGGATTTGTGTGTAGGACCCTG | ACGTTGGATGAAGGAAACAGACCACAGACC | ctCCCTGGAGGCTGAACCCCGTCC | R | Panel I_Exploratory |
| rs6785049 | ACGTTGGATGAGACAGCAGCCACAGTCATC | ACGTTGGATGGCCATCCCATAATCCAGAAG | CCTCCCTCTTCCTCTC | F | Panel I_Exploratory |
| rs2631367 | ACGTTGGATGCCGGGCCTCAGGTGCACTC | ACGTTGGATGCCGCGGCTGGCCTTACATA | ggaaagggaGCCTTGCGGCCCAGGCCCG | R | Panel I_Exploratory |
| rs4795541 | ACGTTGGATGCTAGGATCGCTCCTGCATC | ACGTTGGATGGAGATCCTGGGAGAGGTGC | ACCCCTCGCGGCATCC | R | Panel II_Exploratory |
| rs1800857 | ACGTTGGATGTACAAGAGAATCTGCACCGC | ACGTTGGATGACCCTTGGCTTCGGTTTTAG | CTGGCTGCCACTCTGC | R | Panel II_Exploratory |
| rs1800896 | ACGTTGGATGGAAGTCAGGATTCCATGGAG | ACGTTGGATGCCAAGACAACACTACTAAGG | CCTATCCCTACTTCCCC | R | Panel II_Exploratory |
| rs2569190 | ACGTTGGATGGAGACACAGAACCCTAGATG | ACGTTGGATGCCCCTTCCTTTCCTGGAAAT | AATCCTTCCTGTTACGG | R | Panel II_Exploratory |
| rs35815285 | ACGTTGGATGCAGGCGATTCTCTCCAGAAC | ACGTTGGATGCCGAGACCATCTTCATTGTG | TCTCCAGAACCAGGTGA | R | Panel II_Exploratory |
| rs6311 | ACGTTGGATGTTAGGCTGAAGGGTGAAGAG | ACGTTGGATGACACTGTTGGCTTTGGATGG | ctGAGTGCTGTGAGTGTC | F | Panel II_Exploratory |
| rs4938063 | ACGTTGGATGTCTTGAAAGAGACCCTCTCG | ACGTTGGATGCTATGTGGTCAGCCTGCTAC | ATGACCATGAGGAAGATG | R | Panel II_Exploratory |
| rs352139 | ACGTTGGATGATGATGGCCTGCACCAGGAG | ACGTTGGATGAAGATGCCCATGAAGTGGAG | atTGGCCGGCCCCCAGCTC | R | Panel II_Exploratory |
| rs5030625 | ACGTTGGATGCACTCCAGCTTGGGTGAAAG | ACGTTGGATGAGGGCTTTTACACTTGGCTG | ggTTGGGTGAAAGAGTGAG | F | Panel II_Exploratory |
| rs4570625 | ACGTTGGATGACTCACACATTTGCATGCAC | ACGTTGGATGATCTTATCCCTCCCATCAGC | CATTTGCATGCACAAAATTA | F | Panel II_Exploratory |
| rs5443 | ACGTTGGATGCCCAGACATTGCAGTTGAAG | ACGTTGGATGTCCCACGAGAGCATCATCTG | aCACTGAGGGAGAAGGCCAC | R | Panel II_Exploratory |
| rs28362491 | ACGTTGGATGATGGACCGCATGACTCTATC | ACGTTGGATGTAGGGAAGCCCCCAGGAAG | aGCTGCCTGCGTTCCCCGACC | F | Panel II_Exploratory |

| | | | | | |
|---|---|---|---|---|---|
| rs1800925 | ACGTTGGATGTGACATCAACACCCAACAGG | ACGTTGGATGAGCCATGTCGCCTTTTCCTG | gCTGGAGGACTTCTAGGAAAA | F | Panel II_Exploratory |
| rs1799913 | ACGTTGGATGGTATGTTAAGCACTGCAGCG | ACGTTGGATGACTAGGTTCAGCCAAAAGCG | ctCGTGACAAACTTGTACCTCT | F | Panel II_Exploratory |
| rs4073 | ACGTTGGATGCTGAAGCTCCACAATTTGGT | ACGTTGGATGTGTTCTAACACCTGCCACTC | CCACAATTTGGTGAATTATCAA | R | Panel II_Exploratory |
| rs6313 | ACGTTGGATGTGAGCTCAACTACGAACTCC | ACGTTGGATGTCGACTGTCCAGTTAAATGC | CTCTACAGTAATGACTTTAACTC | F | Panel II_Exploratory |
| rs1800532 | ACGTTGGATGTGGTACCTGGCATGAAATAC | ACGTTGGATGGTTACATTCCCTATGCTCAG | gTTAATTGACAACCTATTAGGTG | F | Panel II_Exploratory |
| rs9858542 | ACGTTGGATGGTTCTCCAAGATCCTCCCTG | ACGTTGGATGAATCATCTCACAGACCCAGC | gggGAGCAAGCTGGCAAACTGAC | F | Panel II_Exploratory |
| rs6314 | ACGTTGGATGCATTCACTCCGTCGCTATTG | ACGTTGGATGGCAAGATGCCAAGACAACAG | cccCTTTAGAAGCCTCTTCAGAAT | R | Panel II_Exploratory |
| rs20541 | ACGTTGGATGGGTCCTGTCTCTGCAAATAA | ACGTTGGATGCCAGTTTGTAAAGGACCTGC | agGCTTTCGAAGTTTCAGTTGAAC | F | Panel II_Exploratory |
| rs4565946 | ACGTTGGATGCTTTGGTTGAAGCAGAGTAG | ACGTTGGATGGTTTTCACTCTGCCATCAGC | GATGGGTTAAATGGCTTAGCTAAG | R | Panel II_Exploratory |
| rs1800795 | ACGTTGGATGGATTGTGCAATGTGACGTCC | ACGTTGGATGAGTGGTTCTGCTTCTTAGCG | TGTGCAATGTGACGTCCTTTAGCAT | R | Panel II_Exploratory |
| rs1800469 | ACGTTGGATGTTCTTACAGGTGTCTGCCTC | ACGTTGGATGAACAAGGTAGGAGAAGAGGG | ggaTCTGCCTCCTGACCCTTCCATCC | F | Panel II_Exploratory |
| rs3813034 | ACGTTGGATGGGGCCCATTACAAATATATG | ACGTTGGATGTGGCAGAGCATGTTGTAGTG | tagGAATTCCCCAAATTTTTCTTACA | R | Panel II_Exploratory |
| rs6897932 | ACGTTGGATGATGGCTACTGAATGCTCACC | ACGTTGGATGAAGATGACCAACAGAGCGAC | atGGGAGATGGATCCTATCTTACTAA | F | Panel II_Exploratory |
| rs1805015 | ACGTTGGATGAACCCTGCTTACCGCAGCTT | ACGTTGGATGTTCTACTTCCTCCAGGTGTC | tgCCGCAGCTTCAGCAAC | F | Panel II_Exploratory |
| rs2631367 | ACGTTGGATGCCGCGGCTGGCCTTACATA | ACGTTGGATGTCCCGCCCCAGCTCCGCCTT | ACGACCAGGGAAGGTT | F | Panel II_Exploratory |
| rs1801275 | ACGTTGGATGAGATCCTCCGCCGAAATGTC | ACGTTGGATGACCCTGCTCCACCGCATGTA | ccCCCCACCAGTGGCTATC | F | Panel II_Exploratory |
| rs1799969 | ACGTTGGATGACTCCCCCACAACTTGTCAG | ACGTTGGATGTCAACCTCTGGTCCCCCAGT | ggagctagttTGGTCTGTTCCCTGGAC | F | Panel II_Exploratory |
| rs4358188 | ACGTTGGATGACTTGTTTGGTTCTCCCCAG | ACGTTGGATGTTGATCTTCATCCAGCCCTC | atgcctggGACCAATTCTGTATCCTCC | F | Panel II_Exploratory |
| rs2243250 | ACGTTGGATGGAATAACAGGCAGACTCTCC | ACGTTGGATGAACTAGGCCTCACCTGATAC | agTCCTACCCCAGCACTGGGG | R | Panel II_Exploratory |
| rs11178997 | ACGTTGGATGGGCACATGTGATATTTTGAC | ACGTTGGATGGGAAGAGTGGAATTGGAATG | ttgATTTTGACACAAGCGTACA | R | Panel II_Exploratory |
| rs10883365 | ACGTTGGATGCTTTCCAAGTGAGAAAGGTC | ACGTTGGATGAGGACGTAGAAAGCTTCCTC | ctcgatcTTCTCAGACGGTTTGAAGGT | F | Panel II_Exploratory |
| rs1800544 | ACGTTGGATGTCCCTTTTCTCCCAAGATCC | ACGTTGGATGCCTGCTGGGAGTTGGCCAT | ctTGCGTTCTGCTCCGTCGGCCC | F | Panel II_Exploratory |
| rs5743836 | ACGTTGGATGTTGCAGTTGACTGTGTAGCC | ACGTTGGATGATAATGGAGGCAAAGGAGGG | cgGATGTGCTGTTCCCTCTGCCTG | R | Panel II_Exploratory |
| rs1982073_ rs1800471#2 | ACGTTGGATGAGTCTTGCAGGTGGATAGTC | ACGTTGGATGTGTTCGCGCTCTCGGCAGT | accGGTGGATAGTCCCGCGGCCGGC | R | Panel II_Exploratory |
| rs1982073_ rs1800471#1 | ACGTTGGATGAGTCTTGCAGGTGGATAGTC | ACGTTGGATGTGTTCGCGCTCTCGGCAGT | gctacgtcCACAGCAGCGGTAGCAGCAGC | R | Panel II_Exploratory |
| rs16944 | ACGTTGGATGCTGTCTGTATTGAGGGTGTG | ACGTTGGATGATTTTCTCCTCAGAGGCTCC | cccACCTTGGGTGCTGTTCTCTGCCTC | F | Panel II_Exploratory |
| rs5498 | ACGTTGGATGACTCACAGAGCACATTCACG | ACGTTGGATGGGAATCAGTGACTGTCACTC | gggACATTCACGGTCACCT | R | Panel II_Exploratory |
| rs16260 | ACGTTGGATGCTAGACCCTAGCAACTCCAG | ACGTTGGATGCCACAGCCAATCAGCAGCG | gggatcatcgaTCCAGGCTAGAGGGCA | F | Panel II_Exploratory |
| rs1800035 | ACGTTGGATGTCGCACCCGCGTGCCACCCA | ACGTTGGATGTTGAGCTGGGTGGGCAGCG | ttaccatcgaCACCGAGCGCAGGCCCAA | F | Panel II_Exploratory |

| | | | | | |
|---|---|---|---|---|---|
| rs36087757 | ACGTTGGATGCCTTTTGAGGTCTCTCTCAC | ACGTTGGATGTCCACTCTCTTTTCAGGAGG | CACCTCCACAGCCACC | F | Panel III_Fine mapping |
| rs187084 | ACGTTGGATGAGCACTTACTATGTGCTGGG | ACGTTGGATGATCGTCTTATTCCCCTGCTG | AAAAGATCACTGCCCT | F | Panel III_Fine mapping |
| rs34466743 | ACGTTGGATGTGACACATCTCTTTGCTCTG | ACGTTGGATGCATCAGCATCAGTCACTTTC | GAGGCTAACGTCGTAA | F | Panel III_Fine mapping |
| rs4537545 | ACGTTGGATGTTTCCCCCTTACTGGTGATG | ACGTTGGATGAGTAAGGACTAGCAAGGAGG | CTGGAAACCCTCCCTGA | F | Panel III_Fine mapping |
| rs5030625 | ACGTTGGATGAGGGCTTTTACACTTGGCTG | ACGTTGGATGCACTCCAGCTTGGGTGAAAG | TCGTTTTGGAGATGGGG | R | Panel III_Fine mapping |
| rs4783566 | ACGTTGGATGTGGACTTGGTTCAGGAAAAA | ACGTTGGATGGGCAGAGCTTGCAGTGAG | tTCCCGTCTGCCATGTTT | F | Panel III_Fine mapping |
| rs1800795 | ACGTTGGATGGATTGTGCAATGTGACGTCC | ACGTTGGATGAGTGGTTCTGCTTCTTAGCG | TGTGACGTCCTTTAGCAT | R | Panel III_Fine mapping |
| rs352144 | ACGTTGGATGTACTGAGTGGGCACTGCTCT | ACGTTGGATGGGCTCTCCTCTCTACAATTC | cccCTCTCCTTCCTGGCTC | F | Panel III_Fine mapping |
| rs7194529 | ACGTTGGATGTTTGAGTGAATGGTGCCCTC | ACGTTGGATGATTCCCTGAGATTCCCCAAC | GGGAAATCATACCTGACAC | R | Panel III_Fine mapping |
| rs8192284 | ACGTTGGATGCTCCAGCAACCAGGAATGTG | ACGTTGGATGCCATATTCTCCTCTTCCTCC | CAGTGGTACTGAAGAAGAA | R | Panel III_Fine mapping |
| rs36215820 | ACGTTGGATGTCACACCAAAGAATCCCACC | ACGTTGGATGATAACATTTCAGGACCCGCC | ccaAGAGGACCACCGTCTCT | F | Panel III_Fine mapping |
| rs1125557 | ACGTTGGATGTCCACTGGGACTCAAGTTTGT | ACGTTGGATGTTCTTTGGAGAGCAACATGC | ACGCTTCAGATGTAGAAATG | R | Panel III_Fine mapping |
| rs1060330 | ACGTTGGATGGCAACAACAGCAAAACTCCG | ACGTTGGATGTAATTCAGGAGGCAGCTTTG | tcAAAAAAAATGCCACCTCAC | F | Panel III_Fine mapping |
| rs2069860 | ACGTTGGATGAGCATCCCTCCACTGCAAAG | ACGTTGGATGCATTTGTGGTTGGGTCAGGG | aaCTTCAGGCAAAGAATCTAG | F | Panel III_Fine mapping |
| rs2276330 | ACGTTGGATGTGTAGTCACCCACCTCTAAG | ACGTTGGATGTATTTTCCTCCCCTGGTCTC | gggCTTGGGCTGGAGAAAGCA | F | Panel III_Fine mapping |
| rs6499203 | ACGTTGGATGGCATCTTACTTTTTAAATTTAGTGTCT | ACGTTGGATGTGTCCACTAGTAGAGAACTGGTTG | cTACGTACCATCATGTACCGAT | R | Panel III_Fine mapping |
| rs10156056 | ACGTTGGATGGGCGTGCTGAATTTAAGGTG | ACGTTGGATGACTGCCAAGTTTATGTCCCC | aTGGGCTGTCCAGAATGCAAGA | F | Panel III_Fine mapping |
| rs352140 | ACGTTGGATGACCTGTCCCACAATAAGCTG | ACGTTGGATGCAAAGGGCTGGCTGTTGTAG | tcGCACTCATTCACGGAGCTACC | R | Panel III_Fine mapping |
| rs2069845 | ACGTTGGATGTTCTCCAGTTTCCCTCTCAC | ACGTTGGATGATGGTGGGTCTATGGAAAGG | tATCTGTTGAAAGACCACTGATC | R | Panel III_Fine mapping |
| rs2069849 | ACGTTGGATGCAACTCATCTCATTCTGCG | ACGTTGGATGAACAACAATCTGAGGTGCCC | gggCTGCGCAGCTTTAAGGAGTT | F | Panel III_Fine mapping |
| rs33964119 | ACGTTGGATGTTGCTGTTTCTTCGGAGGAG | ACGTTGGATGCGCCTCCTTCTTCATCATAG | gggGAGGATGACACCCGGGACAA | F | Panel III_Fine mapping |
| rs4082828 | ACGTTGGATGTGCCAGCAACACAGAGCTAC | ACGTTGGATGAGGAGGTCAAGTCTCTTGTG | CAGCAACACAGAGCTACTTGTTAC | R | Panel III_Fine mapping |
| rs7197744 | ACGTTGGATGCGAACCCAAACATGGAACAC | ACGTTGGATGTAGAGGGCAGAGAAGACTAC | cGCTCTTACTGAAGTTTTTCCAGAG | R | Panel III_Fine mapping |

| | | | | | |
|---|---|---|---|---|---|
| rs2069837 | ACGTTGGATGGTCTCAGTTTCCTTATCTCC | ACGTTGGATGTTAAGTATCTACTGTGTGCC | TTTGAAGATTAGACACAATATTTAT | R | Panel III_Fine mapping |
| rs2140543 | ACGTTGGATGAGCTAAAGAGAGAGATAGGTCCTAATA | ACGTTGGATGGATACATTTGGTCAATAGTGCAGA | gtCTGAAAGTGGGGTATTGATGTTT | R | Panel III_Fine mapping |
| rs35741240 | ACGTTGGATGATTAATCCGGACACTGGTGC | ACGTTGGATGCCCTTACCATTGTCTGTAGC | ccTTTTGAGCACGTGAAGAACAGCAC | F | Panel III_Fine mapping |
| rs3931740 | ACGTTGGATGGAAGTAGGAAGAGCAAGAGC | ACGTTGGATGTGACTCCACACAGCTATTTC | gggGCAAGAGCATGATTTTGAAAAAA | R | Panel III_Fine mapping |
| rs2069827 | ACGTTGGATGGAGTAAATGCCCAACAGAGG | ACGTTGGATGGGTGTTACCTTCACAATCGG | cctCCCAACAGAGGTCACTGTTTTATC | F | Panel III_Fine mapping |
| rs13689 | ACGTTGGATGTAGTCAAGATGTGGCCAGAC | ACGTTGGATGGTTGCTTTGCCCAAGATAGG | atTCAAGATGTGGCCAGACAAAGACAC | R | Panel III_Fine mapping |
| rs2276329 | ACGTTGGATGGGGAGGGAGTGAATATTTGC | ACGTTGGATGAGTAGCATGTGTGGATCCAG | TCTAATCATCATGTCCTATGAAAATTAG | R | Panel III_Fine mapping |
| rs11712164 | ACGTTGGATGACTATAGGTGTGAGCCACTG | ACGTTGGATGTAGGGCCAGACTTGGTTTTC | TGAGATGTAGATCACACAAAAGGACTCA | F | Panel III_Fine mapping |
| rs6969927 | ACGTTGGATGTACCTACTGTAGTACCTGGC | ACGTTGGATGGTCCAGGCTTGGCTAATTTC | tccGAAGTATGTGACTTCCCTTTTCCCTT | F | Panel III_Fine mapping |
| rs2069824 | ACGTTGGATGGAGGATGGCTGGATGGTTTC | ACGTTGGATGAGTGGCTTCGTTTCATGCAG | ttcGTGCTCTCCTGTCTTAAGCAACGTAG | R | Panel III_Fine mapping |
| rs16260_as | ACGTTGGATGCCACAGCCAATCAGCAGCG | ACGTTGGATGCTAGACCCTAGCAACTCCAG | tGCCCGCCCACCCGGCCTCGCATAGACGCG | R | Panel III_Fine mapping |
| rs352143 | ACGTTGGATGTGGAAAGCAAGCACCAGACC | ACGTTGGATGCACCCACCATCTGGATGTAG | aCAGCGGGCACTCCAGCAGCTCAAGCAGAA | F | Panel III_Fine mapping |
| rs10499563 | ACGTTGGATGCCTGGTCTGGCCTGTATAATAA | ACGTTGGATGGGCTTCAGGGAGACTAAACTG | tgggATTTCTTAATTATTATACAAGCACA | R | Panel III_Fine mapping |
| rs4552807 | ACGTTGGATGCTACTTTCAGTACTCTCTCC | ACGTTGGATGATACCCAACCTTCTTCATGC | gggggCTTTAAGGTTTTTTAACAACACAAA | F | Panel III_Fine mapping |
| rs12597188 | ACGTTGGATGTCTTCTCTACACAGTAGCCG | ACGTTGGATGTGGGAAAAGTGAGTCTCTGG | tgTACAACCTCGGTCAGG | R | Panel III_Fine mapping |
| rs16260_s | ACGTTGGATGCTAGACCCTAGCAACTCCAG | ACGTTGGATGCCACAGCCAATCAGCAGCG | cctaTCCAGGCTAGAGGGTCA | F | Panel III_Fine mapping |
| rs10155987 | ACGTTGGATGCTATGAGCCACTTTATTTGCC | ACGTTGGATGCATAGACAAAGCAAGAAGGG | tgTTTTTTGGTGGTCATTTTG | F | Panel III_Fine mapping |
| rs1546762 | ACGTTGGATGACTGCCCGGAAACAGATTAG | ACGTTGGATGGCCTCTGCTGATCAATTTCC | ctGAAACAGATTAGTACAGTACATG | F | Panel III_Fine mapping |
| rs1801023 | ACGTTGGATGATGCCAACATACCTGATGGG | ACGTTGGATGCATTTCTTGGTCTACGCCTG | GCCCCACTGTATTCAG | R | Panel IV_Fine mapping |
| rs16958383 | ACGTTGGATGAAAAAGCTGCCTAGTGGCTG | ACGTTGGATGGCAAAAAGGTTCACCCTCAG | GGGTTCGGTAGGCTCC | R | Panel IV_Fine mapping |
| rs8056633 | ACGTTGGATGGACAGGGTGACTTTCCCAGA | ACGTTGGATGTGAGCCACCACACACAGC | tcaggtatCATGTAAGAAATCAAGAAACTG | R | Panel IV_Fine mapping |
| rs1554606 | ACGTTGGATGGTGCCACTGTGGTGAGATTT | ACGTTGGATGGCAGCCAGAGAGGGAAAAG | TCCTGGGAAAGGTACTC | F | Panel IV_Fine mapping |
| rs12700386 | ACGTTGGATGGTAAACTGAAACAGTGGGTG | ACGTTGGATGCAAAGCAGGACATTTCCCAG | CCCACTGAGCAGACACCA | R | Panel IV_Fine mapping |

| | | | | | |
|---|---|---|---|---|---|
| rs1801026 | ACGTTGGATGTGAGCTCCCTGAACTCCTC | ACGTTGGATGTGCATCACTAACCAGTCTCTTTT | TGGGAAATGCAGAAATCA | F | Panel IV_Fine mapping |
| rs7188750 | ACGTTGGATGTAACAACCCCAAACTGTCCC | ACGTTGGATGAGATCAGAGGCTCTGAACAC | cTCCCACGTTCTCCCATCA | R | Panel IV_Fine mapping |
| rs2056576 | ACGTTGGATGATTTCCCTCTGACACCTGAC | ACGTTGGATGAGAGGCTGTCATGCACATTC | TCAGTAGTTTCACTGGGCC | F | Panel IV_Fine mapping |
| rs7805828 | ACGTTGGATGATCCTTTTCTTACACTCCCC | ACGTTGGATGAGACCATAAGTGGGTCCAAG | TCTTACACTCCCCAAATAGA | F | Panel IV_Fine mapping |
| rs4783689 | ACGTTGGATGGCCATTGTCTATCATTCCAC | ACGTTGGATGCCATTGTCTCAAGTGAAACG | GTGTACACATTATCGAACTC | F | Panel IV_Fine mapping |
| rs2010724 | ACGTTGGATGGCACCTGGAATTCAGGTTGG | ACGTTGGATGAGGCACTTAACCTGCTTTGG | GGAATTCAGGTTGGGGCTCC | R | Panel IV_Fine mapping |
| rs1801552 | ACGTTGGATGACCAAGTGACCACCTTAGAG | ACGTTGGATGCCCAGAATGGCAGGAATTTG | aGTGACTGTGAAGGGGCCGC | F | Panel IV_Fine mapping |
| rs4783573 | ACGTTGGATGTTGCCCCAGGCAGATATTAC | ACGTTGGATGATGTGCAATACCAGCATCTC | GTATTACAGCTTCCTAATGTC | R | Panel IV_Fine mapping |
| rs352163 | ACGTTGGATGGAGTCCACCTCACATTTGTC | ACGTTGGATGATCAGAGGCGATTAGGTGTG | aCTCACATTTGTCTTCACCACT | R | Panel IV_Fine mapping |
| rs3785076 | ACGTTGGATGTTTCAACCCAGCCTGGAAAC | ACGTTGGATGGGCAAGATGGGTTTTAATCG | CATTATTGTCTTTGACTTTTCA | R | Panel IV_Fine mapping |
| rs1800797 | ACGTTGGATGTGGAGACGCCTTGAAGTAAC | ACGTTGGATGGAGTTCTTCTGTGTTCTGGC | AGTAACTGCACGAAATTTGAGG | F | Panel IV_Fine mapping |
| rs7186053 | ACGTTGGATGTTCTGAGGGATCACTTCCAC | ACGTTGGATGCACTGTGCTGGCAAAACATC | GGCATGACTCCCTTTCATTTAAC | F | Panel IV_Fine mapping |
| rs2011779 | ACGTTGGATGGGCTCAGTGGCTCAAACCTA | ACGTTGGATGATGCCTGGCTGATTTTTGG | ccccTTAGTAGAGATGGGTTTC | F | Panel IV_Fine mapping |
| rs3743674 | ACGTTGGATGAATGCGTCCCTCGCAAGTCA | ACGTTGGATGCAGCCATGGGCCCTTGGAG | cTCCCTCGCAAGTCAGGGGATCCG | R | Panel IV_Fine mapping |
| rs9927789 | ACGTTGGATGCAGGGACTTGTAAGTCTAGT | ACGTTGGATGAGATACGCATGTTGTGTGCC | GTTTGGGGATACCACGTATATTGT | F | Panel IV_Fine mapping |
| rs614288 | ACGTTGGATGAAGGGATGGGTCTTGTTCTG | ACGTTGGATGTGGATCTGGCACCTAATGAG | TGTTCTGTAGGCTTTGGGATTAGA | F | Panel IV_Fine mapping |
| rs11717574 | ACGTTGGATGCTCCCAGCTCTTCTCATTAG | ACGTTGGATGTGCTTTGAGCCAACACAACG | aTCTTCTCATTAGCCATTCAGCAGC | R | Panel IV_Fine mapping |
| rs35081782 | ACGTTGGATGATAACATTTCAGGACCCGCC | ACGTTGGATGTCACACCAAAGAATCCCACC | TTCTTCACCGATTGTCTAAACAGAGA | R | Panel IV_Fine mapping |
| rs10216011 | ACGTTGGATGTGAACTTTGGTGTCCCCATC | ACGTTGGATGTATCTGGCACAGTGTATGGC | ATTTGTGTGAATGGATTTGTTCAACA | F | Panel IV_Fine mapping |
| rs8061932 | ACGTTGGATGCTCATTGCCATTCTCTGGAC | ACGTTGGATGGAAAAGGAGTTTGGTCTAGC | ATAGAACAACTGGTCTGCTTTGTTTCT | F | Panel IV_Fine mapping |
| rs3087225 | ACGTTGGATGTTCAGAACATCTTTGGTTTTTACA | ACGTTGGATGCTGAAGTCATGCACGAAGTTT | gGAGAAAAAAAAATAAGCTAAAGCTATG | R | Panel IV_Fine mapping |
| rs7776857 | ACGTTGGATGACTTAGAGCACCTAGCTCAC | ACGTTGGATGTCCTTCTGGTGGATATTCCC | cTTACTGTACAGGGCTCGGCTAGAGCAA | F | Panel IV_Fine mapping |
| rs2069861 | ACGTTGGATGGCAACTTTGAGTGTGTCACG | ACGTTGGATGGCTTTTTTTGGTCGTGGTGG | TGAGTGTGTCACGTGAAGCTTAATATAAA | F | Panel IV_Fine mapping |

| | | | | | |
|---|---|---|---|---|---|
| rs6949149 | ACGTTGGATGTCAGCAAGGTTGGGTCAATC | ACGTTGGATGAGAGGCACAAAAGAGGTCTG | GTCAATCTCACTCCCC | R | Panel IV_Fine mapping |
| rs5743836 | ACGTTGGATGTTGCAGTTGACTGTGTAGCC | ACGTTGGATGATAATGGAGGCAAAGGAGGG | cctaCTGTTCCCTCTGCCTG | R | Panel IV_Fine mapping |
| rs2069840 | ACGTTGGATGCCAGGCAGCAACAAAAAGTG | ACGTTGGATGCTGTCCAAGAATAAACTGCC | cctgctctatTTCATGAGGAGGCCAA | F | Panel IV_Fine mapping |
| rs1800796 | ACGTTGGATGTGGAGACGCCTTGAAGTAAC | ACGTTGGATGGAGTTCTTCTGTGTTCTGGC | tGCAGTTCTACAACAGCC | F | Panel IV_Fine mapping |
| rs3087226 | ACGTTGGATGAGAGTCTCAACCCCCAATAAA | ACGTTGGATGCAGTTCCAGGGCTAAGGATT | ccgttccccctaGGGGAGATAGAGCTTCT | R | Panel IV_Fine mapping |
| rs13447446 | ACGTTGGATGGGACGTCACATTGCACAATC | ACGTTGGATGATCTTTGTTGGAGGGTGAGG | tttgTCAGCCCCACCCGCTCTG | R | Panel IV_Fine mapping |
| rs13335980 | ACGTTGGATGGTGGTCCTGACCTAGGGAAT | ACGTTGGATGAGCCTCCTGAAGTGTTGGA | ccgttccccctaTCGAGCGGCCTAATTTTT | R | Panel IV_Fine mapping |
| rs352139 | ACGTTGGATGAAGATGCCCATGAAGTGGAG | ACGTTGGATGATGATGGCCTGCACCAGGAG | cctgctctatAGTGGAGTGGGTGGAGGT | F | Panel IV_Fine mapping |
| rs1818879 | ACGTTGGATGGGATGCCTCAATAAAGCAGAC | ACGTTGGATGCTGGTCTTCTGACCTCAAGTG | ggTGGGCGCAGTGGCTCAC | R | Panel IV_Fine mapping |
| rs2862231 | ACGTTGGATGGTCGAGCTGTAGTTTCTCTG | ACGTTGGATGTTCTTCCAGTCTTTGACCAC | aCTCTGTAAAAAACTTCGAA | F | Panel IV_Fine mapping |
| rs7186693 | ACGTTGGATGAGCTTCCTAGGATCATGCAG | ACGTTGGATGAACAAGGCAGCCCTGTCTAC | ctGAGGGTGCACTCTCAAAACT | R | Panel IV_Fine mapping |
| rs4783681 | ACGTTGGATGATTACAGGCTTGAGCCACCG | ACGTTGGATGCTCCATACCCTGTGATGTAG | TTGACATAGTGTGGTGATTTT | R | Panel IV_Fine mapping |
| rs445676 | ACGTTGGATGTGGTGCTGGTGATCCTGAG | ACGTTGGATGAGAAGCTGCGCTGACCACTG | cttCTGACGGCCGCCGCTCCCGCTA | F | Panel IV_Fine mapping |
| rs10431923 | ACGTTGGATGTTCTGAGGGATCACTTCCAC | ACGTTGGATGCACTGTGCTGGCAAAACATC | cgggAGGGATCACTTCCACGTGCAGT | F | Panel IV_Fine mapping |

**Supplementary Table 5.2: Association analysis results evaluating susceptibility to PI-IBS of the exploratory SNPs panels in all eligible PI-IBS patients**

| Gene ID | SNP | Variants | Location | Position (dbSNP129) | Minor Allele | Major Allele | Allele frequency Case | Allele frequency Control | Chi2 | P value | OR[1] | CI.low[2] | CI.high[2] | References[3] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **IBS and serotoninergic pathways candidates** | | | | | | | | | | | | | | |
| *SLC6A4* | rs4795541 | 5HTT-LPR / 44-BP INS/DEL / -/A/AGATGCTGGGGGGCTGCAGGGGGGATGCTGGGGGTGCAGGGG/G | 17q11.2 | 25588443-25588485 | Success rate <90% | | | | | | | | | 1-8 |
| *SLC6A4* | rs3813034 | G>T substitution in 3' UTR | 17q11.2 | 25548930 | 2 | 1 | 0.40 | 0.45 | 3.3767 | 0.0661 | 0.802 | 0.634 | 1.015 | 9 |
| *HTR2A* | rs6314 | H452Y | 13q14-q21 | 46307034 | 4 | 2 | 0.09 | 0.09 | 0.0004 | 0.9844 | 1.004 | 0.678 | 1.486 | 9 |
| *HTR2A* | rs6313 | T102C | 13q14-q21 | 46367940 | 4 | 2 | 0.39 | 0.41 | 0.7039 | 0.4015 | 0.904 | 0.715 | 1.144 | 10-12 |
| *HTR2A* | rs6311 | -G1438A | 13q14-q21 | 46369478 | 4 | 2 | 0.39 | 0.41 | 0.4356 | 0.5093 | 0.924 | 0.730 | 1.169 | 10-11 |
| *HTR3A* | rs4938063 | S253N | 11q23.1-q23.2 | 113362502 | MAF[4]<1% | | | | | | | | | 13 |
| *HTR3A* | rs35815285 | R344H | 11q23.1-q23.2 | 113362871 | MAF<1% | | | | | | | | | 13 |
| *ADRA2A* | rs1800544 | -C1291G | 10q24-q26 | 112826492 | 3 | 2 | 0.28 | 0.27 | 0.6454 | 0.4218 | 1.116 | 0.854 | 1.459 | 14 |
| *ADRA2A* | rs1800035 | C753G or K251N | 10q24-q26 | 112828542 | MAF<1% | | | | | | | | | 14 |
| *GNB3* | rs5443 | -C825T | 12p13 | 6825135 | 4 | 2 | 0.34 | 0.32 | 0.9071 | 0.3409 | 1.121 | 0.886 | 1.420 | 12, 15-16 |
| *CCK* | rs1800857 | T779C | 3pter-p21 | 26100208 | 2 | 4 | 0.15 | 0.12 | 2.4180 | 0.1200 | 1.304 | 0.933 | 1.822 | 17 |
| *TGFB1* | rs1800469 | -C509T | 19q13.1 | 46552135 | 4 | 2 | 0.33 | 0.31 | 0.8586 | 0.3541 | 1.121 | 0.880 | 1.429 | 18-20 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *TGFB1* | rs1982073/ rs1800470 | T869C (codon 10) /P10L | 19q13.1 | 46550761 | MAF<1% | | | | | | | | | 18-20 |
| *TGFB1* | rs1800471 | G915C (codon 25) /R25P | 19q13.1 | 46550716 | Success rate <90% | | | | | | | | | 18-20 |
| *TPH1* | rs1800532 | A218C | 11p15.3-p14 | 18004391 | 1 | 2 | 0.39 | 0.39 | 0.0002 | 0.9896 | 1.002 | 0.794 | 1.264 | 12 |
| *TPH1* | rs1799913 | A779C | 11p15.3-p14 | 18003830 | 1 | 2 | 0.39 | 0.39 | 0.0004 | 0.9845 | 0.998 | 0.791 | 1.258 | 12 |
| *TPH2* | rs11178997 | -T473A | 12q21.1 | 70618419 | 1 | 4 | 0.06 | 0.07 | 0.4403 | 0.5070 | 0.850 | 0.526 | 1.374 | 21-24 |
| *TPH2* | rs4570625 | -G703T | 12q21.1 | 70618189 | 4 | 3 | 0.20 | 0.20 | 0.1861 | 0.6662 | 0.938 | 0.703 | 1.253 | 21-24 |
| *TPH2* | rs4565946 | A90G | 12q21.1 | 70623035 | 4 | 2 | 0.43 | 0.43 | 0.0867 | 0.7685 | 1.035 | 0.822 | 1.304 | 21-24 |
| *IL10* | rs1800872 | -A592C | 1q31-32 | 205013029 | 1 | 2 | 0.23 | 0.25 | 0.6293 | 0.4276 | 0.897 | 0.687 | 1.173 | 25-33 |
| *IL10* | rs1800896 | -G1082A | 1q31-32 | 205013519 | 3 | 1 | 0.49 | 0.48 | 0.1408 | 0.7075 | 1.044 | 0.833 | 1.309 | 25-33 |
| **Intestinal eptithelial barrier candidates** | | | | | | | | | | | | | | |
| *SLC22A4* | rs1050152 | L503F | 5q23.3 | 131704218 | 4 | 2 | 0.43 | 0.42 | 0.3136 | 0.5755 | 1.069 | 0.846 | 1.352 | 34 |
| *SLC22A5* | rs2631367 | -G207C | 5q23.3 | 131733356 | 2 | 3 | 0.48 | 0.47 | 0.1875 | 0.6650 | 1.053 | 0.834 | 1.330 | 34 |
| ***CDH1*** | **rs16260** | **-C160A** | **16q22.1** | **67328534** | **1** | **2** | **0.31** | **0.26** | **3.4243** | **0.0642** | **1.258** | **0.986** | **1.605** | **35-37** |
| *CDH1* | rs5030625 | -G347A | 16q22.1 | 67328347 | 7 | 9 | 0.10 | 0.11 | 0.0176 | 0.8945 | 0.976 | 0.677 | 1.406 | 35-37 |
| *ICAM1* | rs5498 | K469E | 19p13.3-p13.2 | 10256682 | 3 | 1 | 0.37 | 0.39 | 0.9795 | 0.3223 | 0.884 | 0.692 | 1.129 | 38-40 |
| *ICAM1* | rs1799969 | G241R | 19p13.3-p13.2 | 10255791 | 1 | 3 | 0.11 | 0.11 | 0.1410 | 0.7073 | 1.076 | 0.735 | 1.573 | 38-40 |
| *DLG5* | rs1248696 | R30Q | 10q23 | 79286610 | 4 | 2 | 0.11 | 0.09 | 0.5929 | 0.4413 | 1.163 | 0.791 | 1.710 | 41 |
| *MYO9B* | rs1545620 | S1011A | 19p13.1 | 17164773 | 2 | 1 | 0.40 | 0.41 | 0.1859 | 0.6664 | 0.952 | 0.759 | 1.193 | 42 |
| *ABCB1* | rs1045642 | C3435T | 7q21.12 | 86976580 | 2 | 4 | 0.43 | 0.44 | 0.2889 | 0.5910 | 0.940 | 0.749 | 1.179 | 43-45 |
| *ABCB1* | rs3789243 | rs3789243 | 7q21.12 | 87058821 | 4 | 2 | 0.46 | 0.47 | 0.2653 | 0.6065 | 0.943 | 0.753 | 1.180 | 43-45 |
| *NR1I2* | rs3814055 | -C25385T | 3q13.33 | 120982724 | 4 | 2 | 0.42 | 0.39 | 0.9933 | 0.3189 | 1.125 | 0.892 | 1.420 | 46 |
| *NR1I2* | rs6785049 | A7635G | 3q13.33 | 121016422 | 3 | 1 | 0.39 | 0.39 | 0.0074 | 0.9316 | 1.010 | 0.805 | 1.268 | 46 |
| *LTF* | rs7645243 /rs9110 | L632L | 3p21.31 | 46455804 | 3 | 1 | 0.25 | 0.28 | 1.0087 | 0.3152 | 0.877 | 0.678 | 1.133 | 47 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *BPI* | rs4358188 | E216K | 20q11.23-q12 | 36380261 | 1 | 3 | 0.47 | 0.49 | 0.4798 | 0.4885 | 0.925 | 0.741 | 1.154 | 48-50 |
| **Innate immune risk factor candidates** | | | | | | | | | | | | | | |
| *NOD2* | rs2066847 | Leu1007insC | 16q12 | 49321279 | 2 | 9 | 0.04 | 0.03 | 0.6229 | 0.4300 | 1.284 | 0.690 | 2.389 | 51-52 |
| *NOD2* | rs2066844 | R702W | 16q12 | 49303426 | 4 | 2 | 0.05 | 0.04 | 1.1193 | 0.2901 | 1.328 | 0.785 | 2.246 | 51-52 |
| *NOD2* | rs2066845 | G908R | 16q12 | 49314040 | 2 | 3 | 0.02 | 0.01 | 1.8388 | 0.1751 | 1.914 | 0.749 | 4.889 | 51-52 |
| *NOD2* | rs2076756 | rs2076756 | 16q12 | 49314381 | 3 | 1 | 0.30 | 0.27 | 1.8035 | 0.1793 | 1.181 | 0.926 | 1.507 | 51-53 |
| *NOD1* | rs6958571 | ND1+32656*1 | 7p14.3 | 30452234 | 2 | 1 | 0.27 | 0.25 | 0.6413 | 0.4232 | 1.109 | 0.861 | 1.427 | 54 |
| *CARD8* | rs2043211 | Cys10Stop | 19q13.3 | 53429517 | 4 | 1 | 0.32 | 0.31 | 0.1287 | 0.7198 | 1.045 | 0.820 | 1.333 | 55 |
| *TLR4* | rs4986790 | D299G | 9q33.1 | 119515122 | 3 | 1 | 0.05 | 0.07 | 2.6605 | 0.1029 | 0.664 | 0.406 | 1.086 | 56-59 |
| *TLR5* | rs5744168 | Arg392Stop or C1174T | 1q41-1q42 | 221351822 | 4 | 2 | 0.05 | 0.05 | 0.3412 | 0.5591 | 1.155 | 0.712 | 1.872 | 60-61 |
| ***TLR9*** | **rs5743836** | **-T1237C** | **3p21.3** | **52235821** | **2** | **4** | **0.13** | **0.18** | **4.7605** | **0.0291** | **0.701** | **0.509** | **0.964** | **62-63** |
| ***TLR9*** | **rs352139** | **2848G/A or P545P** | **3p21.3** | **52233411** | **1** | **3** | **0.48** | **0.41** | **6.3702** | **0.0116** | **1.338** | **1.067** | **1.678** | **62-63** |
| *CD14* | rs2569190 | -T159C | 5q31.1 | 139993099 | 1 | 3 | 0.49 | 0.45 | 1.7293 | 0.1885 | 1.166 | 0.927 | 1.467 | 64-67 |
| *NFKB1* | rs28362491 | -94delATTG | 4q23-q24 | 103641186 | 9 | 7 | 0.38 | 0.38 | 0.0105 | 0.9184 | 0.988 | 0.778 | 1.254 | 68-69 |
| *ATG16L1* | rs2241880 | T300A | 2q37.1 | 233848106 | 4 | 2 | 0.46 | 0.46 | 0.0131 | 0.9089 | 0.987 | 0.785 | 1.241 | 53,70-72 |
| *PTGER4* | rs4495224 | rs4495224 | 5p13.1 | 40513271 | 2 | 1 | 0.33 | 0.33 | 0.1697 | 0.6804 | 1.052 | 0.828 | 1.336 | 53 |
| *PTGER4* | rs4613763 | rs4613763 | 5p13.1 | 40428484 | 2 | 4 | 0.13 | 0.11 | 0.5673 | 0.4513 | 1.143 | 0.807 | 1.618 | 53, 72-73 |
| *PTGER4* | rs9292777 | rs9292777 | 5p13.1 | 40473704 | 2 | 4 | 0.42 | 0.39 | 1.7841 | 0.1817 | 1.171 | 0.929 | 1.478 | 74 |
| *IRGM* | rs4958847 | rs4958847 | 5q33.1 | 150219779 | 1 | 3 | 0.13 | 0.12 | 0.1550 | 0.6938 | 1.073 | 0.757 | 1.520 | 74 |
| *IRGM* | rs13361189 | rs13361189 | 5q33.1 | 150203579 | 2 | 4 | 0.07 | 0.08 | 0.1832 | 0.6686 | 0.909 | 0.586 | 1.409 | 74 |
| *NKX2-3* | rs10883365 | rs10883365 | 10q24.2 | 101277753 | 3 | 1 | 0.50 | 0.47 | 0.7335 | 0.3918 | 1.105 | 0.879 | 1.389 | 71, 74 |
| *NKX2-3* | rs7078219 | rs7078219 | 10q24.2 | 101264354 | 1 | 3 | 0.39 | 0.39 | 0.0010 | 0.9754 | 0.996 | 0.793 | 1.253 | |
| *PTPN2* | rs2542151 | rs2542151 | 18p11.3-p11.2 | 12769946 | 3 | 4 | 0.18 | 0.18 | 0.0705 | 0.7906 | 0.961 | 0.719 | 1.285 | 71, 73-74 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *PTPN2* | rs2847297 | rs2847297 | 18p11.3-p11.2 | 12787693 | 3 | 1 | 0.34 | 0.36 | 0.8593 | 0.3539 | 0.894 | 0.705 | 1.134 | |
| *TNF* | rs1800629 | -G308A | 6p21.3 | 31651009 | 1 | 3 | 0.17 | 0.18 | 0.0000 | 0.9964 | 0.999 | 0.748 | 1.335 | 25-26, 31, 33, 75-78 |
| *TNF* | rs1799724 | -C857T | 6p21.3 | 31650460 | 4 | 2 | 0.10 | 0.09 | 0.0567 | 0.8117 | 1.048 | 0.711 | 1.547 | 25-26, 31, 33, 75-78 |
| *TNF* | rs1800630 | -C863A | 6p21.3 | 31650454 | 1 | 2 | 0.16 | 0.15 | 0.0151 | 0.9021 | 1.019 | 0.751 | 1.384 | 25-26, 31, 33, 75-78 |
| *TNF* | rs1799964 | -T1031C | 6p21.3 | 31650286 | 2 | 4 | 0.21 | 0.19 | 0.1303 | 0.7181 | 1.052 | 0.800 | 1.383 | 25-26, 31, 33, 75-78 |
| *IL1B* | rs1143627 | -C31T | 2q14 | 113310857 | 2 | 4 | 0.37 | 0.33 | 2.3009 | 0.1293 | 1.195 | 0.949 | 1.505 | 25-26, 79-87 |
| *IL1B* | rs1143634 | F105F | 2q14 | 113306860 | 4 | 2 | 0.21 | 0.24 | 2.0883 | 0.1484 | 0.818 | 0.623 | 1.074 | 25-26, 79-87 |
| *IL1B* | rs16944 | -C511T | 2q14 | 113311337 | 1 | 3 | 0.36 | 0.33 | 2.1558 | 0.1420 | 1.191 | 0.943 | 1.505 | 25-26, 79-87 |
| *IL4* | rs2243250 | -C590T | 5q31.1 | 132037052 | 4 | 2 | 0.11 | 0.15 | 3.8724 | 0.0491 | 0.710 | 0.505 | 0.999 | 88-91 |
| *IL4R* | rs1801275 | Q576R | 16p12.1-p11.2 | 27281900 | 3 | 1 | 0.18 | 0.20 | 0.8683 | 0.3514 | 0.873 | 0.656 | 1.162 | 88-93 |
| *IL4R* | rs1805015 | S503P | 16p12.1-p11.2 | 27281680 | 2 | 4 | 0.13 | 0.17 | 2.4594 | 0.1168 | 0.770 | 0.555 | 1.068 | 88-93 |
| *IL6* | **rs1800795** | **-G174C** | **7p21** | **22733169** | **2** | **3** | **0.44** | **0.39** | **4.6568** | **0.0309** | **1.302** | **1.025** | **1.655** | **25-26, 94** |
| *IL6R* | rs4537545 | rs4537545 | 1q21 | 152685502 | 4 | 2 | 0.45 | 0.41 | 1.8718 | 0.1713 | 1.177 | 0.932 | 1.486 | 95 |
| *IL6R* | rs8192284/rs 2228145 | D358A | 1q21 | 152693593 | 3 | 4 | 0.44 | 0.40 | 1.8934 | 0.1688 | 1.178 | 0.933 | 1.488 | 95 |
| *IL7R* | rs6897932 | T244I | 5p13 | 35910331 | 4 | 2 | 0.30 | 0.29 | 0.1099 | 0.7403 | 1.043 | 0.813 | 1.338 | 96-97 |
| *IL8* | rs4073 | -T251A | 4q12-q13 | 74824887 | 1 | 4 | 0.46 | 0.45 | 0.5214 | 0.4703 | 1.086 | 0.868 | 1.360 | 98-100 |
| *IL12B* | rs10045431 | rs10045431 | 5q31.1-q33.1 | 158747110 | 1 | 2 | 0.32 | 0.29 | 2.9657 | 0.0851 | 1.234 | 0.971 | 1.566 | 73-74, 101-102 |
| *IL12B* | rs6887695 | rs6887695 | 5q31.1-q33.1 | 158755222 | 2 | 3 | 0.27 | 0.31 | 2.3305 | 0.1269 | 0.824 | 0.642 | 1.057 | 73-74, 101-102 |
| *IL13* | rs20541 | R130Q | 5q31 | 132023862 | 4 | 2 | 0.18 | 0.20 | 1.6713 | 0.1961 | 0.820 | 0.606 | 1.108 | 103-107 |
| *IL13* | rs1800925 | -C1112T | 5q31 | 132020707 | 4 | 2 | 0.17 | 0.20 | 1.1866 | 0.2760 | 0.852 | 0.638 | 1.137 | 103-107 |
| *IL23R* | rs11209026 | R381Q | 1p31.2 | 67478545 | 1 | 3 | 0.06 | 0.05 | 0.0450 | 0.8321 | 1.051 | 0.662 | 1.669 | 53, 71, 108 |
| 1 Mb region on 3p21 | rs9858542 | rs9858542 | 3p21 | 49676986 | 1 | 3 | 0.29 | 0.29 | 0.0020 | 0.9643 | 1.006 | 0.777 | 1.302 | 71, 74 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 70kb to closest protein-coding gene | rs224136 | rs224136 | 10q21.1 | 64140680 | 4 | 2 | 0.15 | 0.14 | 0.2031 | 0.6522 | 1.076 | 0.783 | 1.477 | 72 |
| 70kb to closest protein-coding gene | rs224135 | rs224135 | 10q21.1 | 64136807 | 1 | 3 | 0.42 | 0.43 | 0.0239 | 0.8771 | 0.982 | 0.781 | 1.235 | 72 |
| 5q region | rs2188962 | rs2188962 | 5q31 | 131798703 | 4 | 2 | 0.42 | 0.42 | 0.0072 | 0.9326 | 1.010 | 0.800 | 1.275 | 73 |

[1]OR: odds ratio; [2] CI: confidence intervals; [3] Refer to the section at the end of this document entitled "References of exploratory and fine mapping SNP panels"; [4] MAF: minor allele frequency.

**Supplementary Table 5.3: Association analysis results evaluating susceptibility to PI-IBS of fine mapping SNPs panels in all eligible PI-IBS patients**

| # | Gene ID | SNP | Comments[1] | Panel ID[2] | Position (dbSNP129) | Minor Alelle | Major Allele | Allele frequency Case | Allele frequency Control | Chi2 | P value | OR[3] | CI.low[4] | CI.high[4] | Reference[5] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *TLR9* | rs11712164 | Tag | Panel III_Fine mapping | 52219090 | 4 | 1 | 0.37 | 0.40 | 1.8666 | 0.1719 | 0.849 | 0.671 | 1.074 | HapMap |
| 2 | *TLR9* | rs352163 | Published-2 | Panel IV_Fine mapping | 52222149 | 3 | 1 | 0.50 | 0.42 | 7.6265 | 0.0058 | 1.371 | 1.096 | 1.715 | 109 |
| 3 | *TLR9* | rs352140 | Published-1 | Panel III_Fine mapping | 52231736 | 3 | 1 | 0.48 | 0.42 | 6.0432 | 0.0140 | 1.324 | 1.059 | 1.656 | 109-113 |
| 4 | *TLR9* | rs352139 | Published-1 | Panel II_Exploratory/ Panel IV_Fine mapping | 52233411 | 1 | 3 | 0.48 | 0.41 | 6.3702 | 0.0116 | 1.338 | 1.067 | 1.678 | 110-112 |
| 5 | *TLR9* | rs5743836 | Published-1 | Panel II_Exploratory/ Panel IV_Fine mapping | 52235821 | 2 | 4 | 0.13 | 0.18 | 4.7605 | 0.0291 | 0.701 | 0.509 | 0.964 | 109-115 |
| 6 | *TLR9* | rs187084 | Published-1 | Panel III_Fine mapping | 52236070 | 2 | 4 | 0.39 | 0.42 | 1.6970 | 0.1927 | 0.856 | 0.678 | 1.081 | 109-112 |
| 7 | *TLR9* | rs352144 | Published-1 | Panel III_Fine mapping | 52236507 | 2 | 1 | 0.06 | 0.05 | 0.0880 | 0.7668 | 1.082 | 0.643 | 1.819 | 111 |
| 8 | *TLR9* | rs352143 | Published-2 | Panel III_Fine mapping | 52239946 | 3 | 1 | 0.19 | 0.24 | 3.6687 | 0.0554 | 0.761 | 0.575 | 1.006 | 109 |
| 9 | *TLR9* | rs11717574 | Tag | Panel IV_Fine mapping | 52243285 | 3 | 1 | 0.14 | 0.18 | 4.5844 | 0.0323 | 0.708 | 0.516 | 0.971 | HapMap |
| 10 | *TLR9* | rs4082828 | Published-2 | Panel III_Fine mapping | 52249781 | 2 | 1 | 0.11 | 0.15 | 5.6959 | 0.0170 | 0.651 | 0.457 | 0.926 | 109 |
| 11 | *TLR9* | rs1060330 | Published-2 | Panel III_Fine mapping | 52263984 | 1 | 3 | 0.51 | 0.44 | 5.7377 | 0.0166 | 1.312 | 1.051 | 1.639 | 109 |
| 12 | *IL6* | rs6949149 | Tag | Panel IV_Fine mapping | 22715681 | 1 | 2 | 0.06 | 0.07 | 0.4783 | 0.4892 | 0.850 | 0.536 | 1.347 | HapMap |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 13 | *IL6* | rs4552807 | HapTag-Predicted reg | Panel III_Fine mapping | 22717543 | 1 | 4 | 0.39 | 0.46 | 6.1200 | 0.0134 | 0.735 | 0.576 | 0.938 | HapMap_forced |
| 14 | *IL6* | rs10216011 | HapTag-Predicted reg | Panel IV_Fine mapping | 22718347 | 3 | 4 | 0.16 | 0.18 | 1.0532 | 0.3048 | 0.851 | 0.625 | 1.158 | HapMap_forced |
| 15 | *IL6* | rs6969927 | Tag | Panel III_Fine mapping | 22719397 | 1 | 2 | 0.23 | 0.28 | 3.5837 | 0.0584 | 0.772 | 0.591 | 1.009 | HapMap |
| 16 | *IL6* | rs10156056 | Tag | Panel III_Fine mapping | 22720612 | 2 | 3 | 0.09 | 0.11 | 0.5716 | 0.4496 | 0.861 | 0.584 | 1.270 | HapMap |
| 17 | *IL6* | rs7776857 | Tag | Panel IV_Fine mapping | 22721292 | 3 | 4 | 0.36 | 0.33 | 1.1523 | 0.2831 | 1.140 | 0.897 | 1.450 | HapMap |
| 18 | *IL6* | rs10155987 | Tag | Panel III_Fine mapping | 22723060 | 1 | 3 | 0.08 | 0.09 | 0.5875 | 0.4434 | 0.850 | 0.562 | 1.287 | HapMap |
| 19 | *IL6* | rs1546762 | Regulatory | Panel III_Fine mapping | 22724027 | 2 | 4 | 0.51 | 0.46 | 2.8800 | 0.0897 | 1.224 | 0.969 | 1.547 | UCSC |
| 20 | *IL6* | rs7805828 | HapTag-Predicted reg | Panel IV_Fine mapping | 22725086 | 1 | 3 | 0.41 | 0.44 | 0.8817 | 0.3477 | 0.894 | 0.708 | 1.129 | HapMap_forced |
| 21 | *IL6* | rs10499563 | Tag | Panel III_Fine mapping | 22727012 | 3 | 1 | 0.22 | 0.25 | 1.9948 | 0.1578 | 0.823 | 0.627 | 1.079 | HapMap |
| 22 | *IL6* | rs2056576 | Tag | Panel IV_Fine mapping | 22727726 | 4 | 2 | 0.30 | 0.33 | 1.1153 | 0.2909 | 0.876 | 0.685 | 1.120 | HapMap |
| 23 | *IL6* | rs12700386 | Tag | Panel IV_Fine mapping | 22729533 | 2 | 3 | 0.20 | 0.19 | 0.4461 | 0.5042 | 1.103 | 0.828 | 1.469 | HapMap |
| 24 | *IL6* | rs2069824 | HapTag-Predicted reg | Panel III_Fine mapping | 22731756 | 3 | 1 | 0.07 | 0.08 | 1.5710 | 0.2101 | 0.754 | 0.485 | 1.173 | HapMap_forced |
| 26 | *IL6* | rs2069827 | Published/IL-6 level | Panel III_Fine mapping | 22731980 | 4 | 3 | 0.06 | 0.09 | 3.4858 | 0.0619 | 0.644 | 0.406 | 1.022 | 116-118 |
| 27 | *IL6* | rs1800797 | Published/IL-6 level | Panel IV_Fine mapping | 22732745 | 1 | 3 | 0.44 | 0.39 | 3.7332 | 0.0533 | 1.269 | 0.997 | 1.616 | 117, 119-121 |
| 28 | *IL6* | rs1800796 | Published/IL-6 level | Panel IV_Fine mapping | 22732770 | 2 | 3 | 0.06 | 0.05 | 0.1011 | 0.7505 | 1.084 | 0.660 | 1.781 | 119-121 |
| 29 | *IL6* | rs1800795 | Published/IL-6 level | Panel II_Exploratory/ Panel IV_Fine mapping | 22733169 | 2 | 3 | 0.44 | 0.39 | 4.6568 | 0.0309 | 1.302 | 1.025 | 1.655 | 116-117, 119-121 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | *IL6* | rs2069837 | Tag | Panel III_Fine mapping | 22734551 | 2 | 4 | 0.07 | 0.09 | 1.2807 | 0.2578 | 0.775 | 0.499 | 1.205 | HapMap |
| 31 | *IL6* | rs2069840 | Published | Panel IV_Fine mapping | 22735096 | 3 | 2 | 0.34 | 0.37 | 0.7381 | 0.3903 | 0.900 | 0.707 | 1.145 | 118 |
| **32** | ***IL6*** | **rs1554606** | **Published** | **Panel IV_Fine mapping** | **22735231** | **4** | **3** | **0.46** | **0.41** | **4.4515** | **0.0349** | **1.292** | **1.018** | **1.639** | **117-118** |
| **33** | ***IL6*** | **rs2069845** | **Published** | **Panel III_Fine mapping** | **22736673** | **2** | **4** | **0.47** | **0.41** | **5.1106** | **0.0238** | **1.317** | **1.037** | **1.672** | **118** |
| 34 | *IL6* | rs2069849 | Published/coding | Panel III_Fine mapping | 22737680 | 4 | 2 | 0.02 | 0.02 | 0.8559 | 0.3549 | 1.465 | 0.652 | 3.288 | 117-118 |
| **35** | ***IL6*** | **rs2069861** | **Published** | **Panel IV_Fine mapping** | **22738178** | **4** | **2** | **0.14** | **0.09** | **7.0603** | **0.0079** | **1.608** | **1.133** | **2.283** | **117** |
| 36 | *IL6* | rs2140543 | Tag | Panel III_Fine mapping | 22744729 | 3 | 2 | 0.30 | 0.30 | 0.2180 | 0.6406 | 0.939 | 0.721 | 1.223 | HapMap |
| 37 | *CDH1* | rs7194529 | HapTag-Predicted reg | Panel III_Fine mapping | 67295877 | 1 | 4 | 0.10 | 0.11 | 0.1458 | 0.7026 | 0.931 | 0.645 | 1.344 | HapMap-UCSC |
| **38** | ***CDH1*** | **rs7186693** | **HapTag-Predicted reg** | **Panel IV_Fine mapping** | **67303341** | **4** | **2** | **0.54** | **0.49** | **4.1724** | **0.0411** | **1.266** | **1.010** | **1.589** | **HapMap-UCSC** |
| 39 | *CDH1* | rs5030625 | Published | Panel II_Exploratory/ Panel III_Fine mapping | 67328347 | 7 | 9 | 0.10 | 0.11 | 0.0176 | 0.8945 | 0.976 | 0.677 | 1.406 | 122-123 |
| 40 | *CDH1* | rs16260 | Published | Panel II_Exploratory/ Panel III_Fine mapping | 67328534 | 1 | 2 | 0.31 | 0.26 | 3.4243 | 0.0642 | 1.258 | 0.986 | 1.605 | 122-127 |
| 41 | *CDH1* | rs3743674 | Published/Regulatory (intron 1junction) | Panel IV_Fine mapping | 67328872 | 3 | 1 | 0.10 | 0.10 | 0.0076 | 0.9305 | 1.017 | 0.700 | 1.477 | 126 |
| 42 | *CDH1* | rs2862231 | Published | Panel IV_Fine mapping | 67331708 | 4 | 2 | 0.10 | 0.11 | 0.0254 | 0.8733 | 0.971 | 0.672 | 1.401 | 127 |
| 43 | *CDH1* | rs3931740 | Tag | Panel III_Fine mapping | 67360722 | 1 | 2 | 0.07 | 0.06 | 0.5493 | 0.4586 | 1.181 | 0.761 | 1.834 | HapMap |
| 44 | *CDH1* | rs4783681 | Published/Predicted reg | Panel IV_Fine mapping | 67364008 | 4 | 3 | 0.32 | 0.26 | 3.7747 | 0.0520 | 1.279 | 0.998 | 1.639 | 127 |
| 45 | *CDH1* | rs1125557 | Published | Panel III_Fine mapping | 67367099 | 2 | 4 | 0.43 | 0.39 | 2.3574 | 0.1247 | 1.197 | 0.952 | 1.505 | 127 |
| 46 | *CDH1* | rs12597188 | Tag | Panel III_Fine mapping | 67372326 | 4 | 2 | 0.37 | 0.33 | 1.7516 | 0.1857 | 1.178 | 0.924 | 1.500 | HapMap |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 47 | *CDH1* | rs2010724 | Published | Panel IV_Fine mapping | 67389914 | 2 | 4 | 0.36 | 0.30 | 4.6414 | 0.0312 | 1.296 | 1.024 | 1.641 | 127 |
| 48 | *CDH1* | rs10431923 | Tag | Panel IV_Fine mapping | 67396763 | 3 | 4 | 0.49 | 0.44 | 4.2271 | 0.0398 | 1.270 | 1.011 | 1.595 | HapMap |
| 49 | *CDH1* | rs7186053 | Tag | Panel IV_Fine mapping | 67396793 | 1 | 3 | 0.44 | 0.38 | 5.2584 | 0.0218 | 1.302 | 1.039 | 1.631 | HapMap |
| 50 | *CDH1* | rs4783573 | Tag | Panel IV_Fine mapping | 67398088 | 2 | 4 | 0.29 | 0.33 | 1.9813 | 0.1593 | 0.840 | 0.660 | 1.071 | HapMap |
| 51 | *CDH1* | rs7188750 | Tag | Panel IV_Fine mapping | 67400395 | 4 | 2 | 0.16 | 0.17 | 0.0617 | 0.8039 | 1.039 | 0.767 | 1.407 | HapMap |
| 52 | *CDH1* | rs3785076 | Tag | Panel IV_Fine mapping | 67408283 | 2 | 4 | 0.03 | 0.04 | 0.3507 | 0.5537 | 0.820 | 0.425 | 1.581 | HapMap |
| 53 | *CDH1* | rs4783689 | Tag | Panel IV_Fine mapping | 67411171 | 4 | 2 | 0.43 | 0.41 | 0.4223 | 0.5158 | 1.077 | 0.861 | 1.346 | HapMap |
| 54 | *CDH1* | rs16958383 | Tag | Panel IV_Fine mapping | 67414500 | 4 | 2 | 0.16 | 0.16 | 0.3496 | 0.5544 | 1.095 | 0.810 | 1.482 | HapMap |
| 55 | *CDH1* | rs2276330 | Regulatory/ Predicted reg | Panel III_Fine mapping | 67414789 | 3 | 1 | 0.14 | 0.15 | 0.4028 | 0.5257 | 1.107 | 0.809 | 1.514 | UCSC |
| 56 | *CDH1* | rs1801552 | Coding | Panel IV_Fine mapping | 67414941 | 4 | 2 | 0.36 | 0.36 | 0.5138 | 0.4735 | 0.916 | 0.720 | 1.165 | dbSNP |
| 57 | *CDH1* | rs33964119 | Coding | Panel III_Fine mapping | 67419665 | 4 | 2 | 0.02 | 0.03 | 0.4966 | 0.4810 | 0.765 | 0.363 | 1.611 | dbSNP |
| 58 | *CDH1* | rs8061932 | Tag | Panel IV_Fine mapping | 67420097 | 2 | 4 | 0.16 | 0.17 | 0.0007 | 0.9792 | 1.004 | 0.743 | 1.357 | HapMap |
| 59 | *CDH1* | rs9927789 | Tag | Panel IV_Fine mapping | 67420415 | 2 | 1 | 0.08 | 0.10 | 1.7559 | 0.1851 | 0.754 | 0.497 | 1.145 | HapMap |
| 60 | *CDH1* | rs2276329 | Published | Panel III_Fine mapping | 67420814 | 2 | 4 | 0.07 | 0.08 | 0.0382 | 0.8451 | 0.957 | 0.617 | 1.485 | 127 |
| 61 | *CDH1* | rs1801026 | Published | Panel IV_Fine mapping | 67424956 | 4 | 2 | 0.15 | 0.16 | 0.0002 | 0.9888 | 1.002 | 0.734 | 1.368 | 127 |
| 62 | *CDH1* | rs13689 | HapTag-Predicted reg | Panel III_Fine mapping | 67426022 | 3 | 1 | 0.20 | 0.21 | 0.0381 | 0.8453 | 1.029 | 0.772 | 1.373 | HapMap-UCSC |
| 63 | *CDH1* | rs7197744 | HapTag-Predicted reg | Panel III_Fine mapping | 67438878 | 1 | 2 | 0.13 | 0.13 | 0.3329 | 0.5640 | 1.102 | 0.792 | 1.533 | HapMap-UCSC |
| 64 | *CDH1* | rs6499203 | Tag | Panel III_Fine mapping | 67442893 | 2 | 4 | 0.23 | 0.25 | 0.0964 | 0.7563 | 0.958 | 0.732 | 1.254 | HapMap |

[1]"Tag": SNP selected by tagging algorithm, "published": SNP previously reported in the literature (see reference column), "predicted reg": SNP located in a predicted regulatory region based on the computational method called ESPERR (Evolutionary and Sequence Pattern Extraction through Reduced Representations) that is also available from the UCSC genome browser (http://genome.ucsc.edu/) under the name "7X Reg Potential" within the "Expression and Regulation" track group, "coding": SNP located in coding region of the candidate gene; [2] Genotyping panel ID referred to in the last column of Supplementary Table 1; [3]OR: odds ratio; [4] CI: confidence intervals; [5] Refer to the section at the end of this document entitled "References of exploratory and fine mapping SNP panels", "HapMap": SNP taken from the HapMap database (http://www.hapmap.org/), "HapMap_forced": SNP taken from the HapMap database that was forced into the tagging SNP design, "UCSC": SNP located in predicted regulatory region observed using the UCSC genome browser, "HapMap-UCSC": SNP present in the HapMap database located in predicted regulatory region (see UCSC), "dbSNP": SNP located in coding region of candidate gene taken from the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/).

**Supplementary Table 5.4: Association analysis results evaluating susceptibility to PI-IBS of the fine mapping SNPs panels in all confirmed eligible PI-IBS patients**

| # | Gene ID | SNP | Panel ID[1] | Position (dbSNP129) | Minor Allele | Major Allele | Allele Frequency Case | Allele Frequency Control | Chi2 | P value | OR[2] | CI.low[3] | CI.high[3] | Reference[4] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *TLR9* | rs11712164 | Panel III_Fine mapping | 52219090 | 4 | 1 | 0.35 | 0.36 | 0.0892 | 0.7652 | 0.942 | 0.638 | 1.392 | HapMap |
| 2 | *TLR9* | rs352163 | Panel IV_Fine mapping | 52222149 | 3 | 1 | 0.51 | 0.46 | 1.2870 | 0.2566 | 1.242 | 0.854 | 1.804 | 109 |
| 3 | *TLR9* | rs352140 | Panel III_Fine mapping | 52231736 | 3 | 1 | 0.49 | 0.46 | 0.4691 | 0.4934 | 1.137 | 0.787 | 1.645 | 109-113 |
| 4 | *TLR9* | rs352139 | Panel II_Exploratory/ Panel IV_Fine mapping | 52233411 | 1 | 3 | 0.48 | 0.44 | 0.8231 | 0.3643 | 1.187 | 0.820 | 1.719 | 110-112 |
| 5 | *TLR9* | rs5743836 | Panel II_Exploratory/ Panel IV_Fine mapping | 52235821 | 2 | 4 | 0.13 | 0.21 | 3.6787 | 0.0551 | 0.600 | 0.356 | 1.011 | 109-115 |
| 6 | *TLR9* | rs187084 | Panel III_Fine mapping | 52236070 | 2 | 4 | 0.39 | 0.36 | 0.1659 | 0.6837 | 1.084 | 0.736 | 1.595 | 109-112 |
| 7 | *TLR9* | rs352144 | Panel III_Fine mapping | 52236507 | 2 | 1 | 0.06 | 0.04 | 1.3806 | 0.2400 | 1.684 | 0.706 | 4.017 | 111 |
| 8 | *TLR9* | rs352143 | Panel III_Fine mapping | 52239946 | 3 | 1 | 0.20 | 0.25 | 1.6853 | 0.1942 | 0.738 | 0.467 | 1.167 | 109 |
| 9 | *TLR9* | rs11717574 | Panel IV_Fine mapping | 52243285 | 3 | 1 | 0.14 | 0.22 | 3.3307 | 0.0680 | 0.622 | 0.373 | 1.036 | HapMap |
| 10 | *TLR9* | rs4082828 | Panel III_Fine mapping | 52249781 | 2 | 1 | 0.11 | 0.18 | 4.0037 | 0.0454 | 0.558 | 0.315 | 0.988 | 109 |
| 11 | *TLR9* | rs1060330 | Panel III_Fine mapping | 52263984 | 1 | 3 | 0.53 | 0.47 | 1.3725 | 0.2414 | 1.246 | 0.862 | 1.800 | 109 |
| 12 | *IL6* | rs6949149 | Panel IV_Fine mapping | 22715681 | 1 | 2 | 0.08 | 0.07 | 0.1870 | 0.6654 | 1.160 | 0.592 | 2.274 | HapMap |
| 13 | *IL6* | rs4552807 | Panel III_Fine mapping | 22717543 | 1 | 4 | 0.40 | 0.48 | 2.9041 | 0.0884 | 0.696 | 0.459 | 1.056 | HapMap_forced |
| 14 | *IL6* | rs10216011 | Panel IV_Fine mapping | 22718347 | 3 | 4 | 0.18 | 0.17 | 0.1656 | 0.6840 | 1.103 | 0.689 | 1.766 | HapMap_forced |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | *IL6* | rs6969927 | Panel III_Fine mapping | 22719397 | 1 | 2 | 0.21 | 0.30 | 4.4927 | 0.0340 | 0.617 | 0.394 | 0.964 | HapMap |
| 16 | *IL6* | rs10156056 | Panel III_Fine mapping | 22720612 | 2 | 3 | 0.10 | 0.10 | 0.0018 | 0.9658 | 0.987 | 0.538 | 1.810 | HapMap |
| 17 | *IL6* | rs7776857 | Panel IV_Fine mapping | 22721292 | 3 | 4 | 0.39 | 0.32 | 1.7883 | 0.1811 | 1.307 | 0.883 | 1.935 | HapMap |
| 18 | *IL6* | rs10155987 | Panel III_Fine mapping | 22723060 | 1 | 3 | 0.09 | 0.07 | 0.6371 | 0.4248 | 1.323 | 0.666 | 2.628 | HapMap |
| 19 | *IL6* | rs1546762 | Panel III_Fine mapping | 22724027 | 2 | 4 | 0.56 | 0.46 | 4.2160 | 0.0400 | 1.497 | 1.018 | 2.199 | UCSC |
| 20 | *IL6* | rs7805828 | Panel IV_Fine mapping | 22725086 | 1 | 3 | 0.35 | 0.46 | 4.9569 | 0.0260 | 0.645 | 0.439 | 0.949 | HapMap_forced |
| 21 | *IL6* | rs10499563 | Panel III_Fine mapping | 22727012 | 3 | 1 | 0.18 | 0.25 | 3.2914 | 0.0696 | 0.646 | 0.403 | 1.036 | HapMap |
| 22 | *IL6* | rs2056576 | Panel IV_Fine mapping | 22727726 | 4 | 2 | 0.28 | 0.35 | 1.9124 | 0.1667 | 0.755 | 0.507 | 1.124 | HapMap |
| 23 | *IL6* | rs12700386 | Panel IV_Fine mapping | 22729533 | 2 | 3 | 0.17 | 0.20 | 0.5868 | 0.4437 | 0.832 | 0.520 | 1.331 | HapMap |
| 24 | *IL6* | rs2069824 | Panel III_Fine mapping | 22731756 | 3 | 1 | 0.07 | 0.08 | 0.0789 | 0.7788 | 0.907 | 0.460 | 1.790 | HapMap_forced |
| 26 | *IL6* | rs2069827 | Panel III_Fine mapping | 22731980 | 4 | 3 | 0.07 | 0.09 | 0.8318 | 0.3618 | 0.706 | 0.334 | 1.491 | 116-118 |
| 27 | *IL6* | rs1800797 | Panel IV_Fine mapping | 22732745 | 1 | 3 | 0.47 | 0.39 | 2.5990 | 0.1069 | 1.386 | 0.932 | 2.062 | 117, 119-121 |
| 28 | *IL6* | rs1800796 | Panel IV_Fine mapping | 22732770 | 2 | 3 | 0.07 | 0.04 | 1.1836 | 0.2766 | 1.511 | 0.718 | 3.181 | 119-121 |
| 29 | *IL6* | rs1800795 | Panel II_Exploratory/ Panel IV_Fine mapping | 22733169 | 2 | 3 | 0.48 | 0.40 | 3.1839 | 0.0744 | 1.433 | 0.965 | 2.129 | 116-117, 119-121 |
| 30 | *IL6* | rs2069837 | Panel III_Fine mapping | 22734551 | 2 | 4 | 0.05 | 0.09 | 1.7227 | 0.1894 | 0.600 | 0.280 | 1.287 | HapMap |
| 31 | *IL6* | rs2069840 | Panel IV_Fine mapping | 22735096 | 3 | 2 | 0.29 | 0.38 | 3.7695 | 0.0522 | 0.663 | 0.438 | 1.004 | 118 |
| 32 | *IL6* | rs1554606 | Panel IV_Fine mapping | 22735231 | 4 | 3 | 0.51 | 0.41 | 4.5558 | 0.0328 | 1.534 | 1.036 | 2.274 | 117-118 |
| 33 | *IL6* | rs2069845 | Panel III_Fine mapping | 22736673 | 2 | 4 | 0.51 | 0.41 | 4.8361 | 0.0279 | 1.560 | 1.050 | 2.318 | 118 |
| 34 | *IL6* | rs2069849 | Panel III_Fine mapping | 22737680 | 4 | 2 | 0.02 | 0.01 | 1.7593 | 0.1847 | 2.798 | 0.612 | 12.803 | 117-118 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 35 | *IL6* | rs2069861 | Panel IV_Fine mapping | 22738178 | 4 | 2 | 0.20 | 0.07 | 13.3541 | 0.0003 | 2.840 | 1.622 | 4.970 | 117 |
| 36 | *IL6* | rs2140543 | Panel III_Fine mapping | 22744729 | 3 | 2 | 0.27 | 0.30 | 0.7465 | 0.3876 | 0.832 | 0.549 | 1.262 | HapMap |
| 37 | *CDH1* | rs7194529 | Panel III_Fine mapping | 67295877 | 1 | 4 | 0.10 | 0.13 | 0.9956 | 0.3184 | 0.736 | 0.403 | 1.344 | HapMap-UCSC |
| 38 | *CDH1* | rs7186693 | Panel IV_Fine mapping | 67303341 | 4 | 2 | 0.58 | 0.47 | 4.7881 | 0.0287 | 1.506 | 1.044 | 2.175 | HapMap-UCSC |
| 39 | *CDH1* | rs5030625 | Panel II_Exploratory/ Panel III_Fine mapping | 67328347 | 7 | 9 | 0.10 | 0.13 | 0.8816 | 0.3478 | 0.748 | 0.408 | 1.371 | 122-123 |
| 40 | *CDH1* | rs16260 | Panel II_Exploratory/ Panel III_Fine mapping | 67328534 | 1 | 2 | 0.35 | 0.21 | 9.7864 | 0.0018 | 1.942 | 1.281 | 2.943 | 122-127 |
| 41 | *CDH1* | rs3743674 | Panel IV_Fine mapping | 67328872 | 3 | 1 | 0.10 | 0.13 | 0.8615 | 0.3533 | 0.752 | 0.411 | 1.374 | 126 |
| 42 | *CDH1* | rs2862231 | Panel IV_Fine mapping | 67331708 | 4 | 2 | 0.10 | 0.13 | 0.9450 | 0.3310 | 0.742 | 0.406 | 1.354 | 127 |
| 43 | *CDH1* | rs3931740 | Panel III_Fine mapping | 67360722 | 1 | 2 | 0.06 | 0.07 | 0.0868 | 0.7683 | 0.896 | 0.432 | 1.860 | HapMap |
| 44 | *CDH1* | rs4783681 | Panel IV_Fine mapping | 67364008 | 4 | 3 | 0.36 | 0.21 | 11.7007 | 0.0006 | 2.071 | 1.365 | 3.143 | 127 |
| 45 | *CDH1* | rs1125557 | Panel III_Fine mapping | 67367099 | 2 | 4 | 0.49 | 0.34 | 8.7607 | 0.0031 | 1.751 | 1.208 | 2.537 | 127 |
| 46 | *CDH1* | rs12597188 | Panel III_Fine mapping | 67372326 | 4 | 2 | 0.41 | 0.29 | 7.4728 | 0.0063 | 1.777 | 1.177 | 2.684 | HapMap |
| 47 | *CDH1* | rs2010724 | Panel IV_Fine mapping | 67389914 | 2 | 4 | 0.45 | 0.25 | 15.4575 | 0.0001 | 2.142 | 1.465 | 3.131 | 127 |
| 48 | *CDH1* | rs10431923 | Panel IV_Fine mapping | 67396763 | 3 | 4 | 0.55 | 0.39 | 10.0041 | 0.0016 | 1.813 | 1.254 | 2.622 | HapMap |
| 49 | *CDH1* | rs7186053 | Panel IV_Fine mapping | 67396793 | 1 | 3 | 0.48 | 0.32 | 10.7365 | 0.0011 | 1.870 | 1.286 | 2.720 | HapMap |
| 50 | *CDH1* | rs4783573 | Panel IV_Fine mapping | 67398088 | 2 | 4 | 0.28 | 0.37 | 4.0432 | 0.0444 | 0.657 | 0.436 | 0.989 | HapMap |
| 51 | *CDH1* | rs7188750 | Panel IV_Fine mapping | 67400395 | 4 | 2 | 0.12 | 0.21 | 5.1217 | 0.0236 | 0.544 | 0.322 | 0.922 | HapMap |
| 52 | *CDH1* | rs3785076 | Panel IV_Fine mapping | 67408283 | 2 | 4 | 0.02 | 0.04 | 0.4562 | 0.4994 | 0.669 | 0.209 | 2.145 | HapMap |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 53 | *CDH1* | rs4783689 | Panel IV_Fine mapping | 67411171 | 4 | 2 | 0.48 | 0.41 | 1.6467 | 0.1994 | 1.271 | 0.881 | 1.832 | HapMap |
| 54 | *CDH1* | rs16958383 | Panel IV_Fine mapping | 67414500 | 4 | 2 | 0.14 | 0.20 | 1.8088 | 0.1787 | 0.712 | 0.435 | 1.168 | HapMap |
| 55 | *CDH1* | rs2276330 | Panel III_Fine mapping | 67414789 | 3 | 1 | 0.12 | 0.20 | 4.0619 | 0.0439 | 0.587 | 0.350 | 0.985 | UCSC |
| 56 | *CDH1* | rs1801552 | Panel IV_Fine mapping | 67414941 | 4 | 2 | 0.33 | 0.32 | 0.0379 | 0.8456 | 1.039 | 0.705 | 1.531 | dbSNP |
| 57 | *CDH1* | rs33964119 | Panel III_Fine mapping | 67419665 | 4 | 2 | 0.01 | 0.03 | 1.4447 | 0.2294 | 0.386 | 0.082 | 1.823 | dbSNP |
| 58 | *CDH1* | rs8061932 | Panel IV_Fine mapping | 67420097 | 2 | 4 | 0.14 | 0.23 | 4.7476 | 0.0293 | 0.572 | 0.346 | 0.945 | HapMap |
| 59 | *CDH1* | rs9927789 | Panel IV_Fine mapping | 67420415 | 2 | 1 | 0.07 | 0.10 | 1.0587 | 0.3035 | 0.697 | 0.350 | 1.387 | HapMap |
| 60 | *CDH1* | rs2276329 | Panel III_Fine mapping | 67420814 | 2 | 4 | 0.06 | 0.09 | 1.5908 | 0.2072 | 0.610 | 0.283 | 1.314 | 127 |
| 61 | *CDH1* | rs1801026 | Panel IV_Fine mapping | 67424956 | 4 | 2 | 0.11 | 0.21 | 6.9610 | 0.0083 | 0.472 | 0.270 | 0.825 | 127 |
| 62 | *CDH1* | rs13689 | Panel III_Fine mapping | 67426022 | 3 | 1 | 0.14 | 0.27 | 8.4281 | 0.0037 | 0.469 | 0.281 | 0.782 | HapMap-UCSC |
| 63 | *CDH1* | rs7197744 | Panel III_Fine mapping | 67438878 | 1 | 2 | 0.10 | 0.18 | 4.8179 | 0.0282 | 0.524 | 0.294 | 0.933 | HapMap-UCSC |
| 64 | *CDH1* | rs6499203 | Panel III_Fine mapping | 67442893 | 2 | 4 | 0.17 | 0.31 | 8.5384 | 0.0035 | 0.494 | 0.308 | 0.793 | HapMap |

[1] Genotyping panel ID referred to in Supplementary Table 1; [2]OR: odds ratio; [3] CI: confidence intervals; [4] Refer to the section at the end of this document entitled "References of exploratory and fine mapping SNP panels", "HapMap": SNP taken from the HapMap database (http://www.hapmap.org/), "HapMap_forced": SNP taken from the HapMap database that was forced into the tagging SNP design, "UCSC": SNP located in predicted regulatory region observed from the UCSC genome browser, "HapMap-UCSC": SNP present in the HapMap database located in predicted regulatory region (see UCSC), "dbSNP": SNP located in coding region of candidate gene taken from the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/).

### 5.2.11 Supplementary References: exploratory and fine mapping candidates' selection


1. Park JM, Choi MG, Park JA, Oh JH, Cho YK, Lee IS, Kim SW, Choi KY, Chung IS. Serotonin transporter gene polymorphism and irritable bowel syndrome. Neurogastroenterol Motil 2006; 18 (11):995-1000.

2. Mawe GM, Coates MD, Moses PL. Review article: intestinal serotonin signalling in irritable bowel syndrome. Aliment Pharmacol Ther 2006; 23(8): 1067-1076.

3. Yeo A, Boyd P, Lumsden S, Saunders T, Handley A, Stubbins M, Knaggs A, Asquith S, Taylor I, Bahari B, Crocker N, Rallan R, Varsani S, Montgomery D, Alpers DH, Dukes GE, Purvis I, Hicks GA. Association between a functional polymorphism in the serotonin transporter gene and diarrhoea predominant irritable bowel syndrome in women. Gut 2004; 53 (10): 1452-1458.

4. Scherl E, Frissora CL. Irritable bowel syndrome genophenomics: correlation serotonin-transporter polymorphisms and alosetron response. Pharmacogenomics J 2003; 3 (2): 64-66.

5. Pata C, Erdal ME, Derici E, Yazar A, Kanik A, Ulu O. Serotonin transporter gene polymorphism in irritable bowel syndrome. Am J Gastroenterol 2002; 97 (7): 1780-1784.

6. Camilleri M, Atanasova E, Carlson PJ, Ahmad U, Kim HJ, Viramontes BE, McKinzie S, Urrutia R. Serotonin-transporter polymorphism pharmacogenetics in diarrhea-predominant irritable bowel syndrome. Gastroenterology 2002; 123 (2): 425-432.

7. Li Y, Nie Y, Xie J, Tang W, Liang P, Sha W, Yang H, Zhou Y. The Association of Serotonin Transporter Genetic Polymorphisms and Irritable Bowel Syndrome and Its Influence on Tegaserod Treatment in Chinese Patients. Dig Dis Sci 2007; 52: 2942-2949.

8. Geeraerts B, van Oudenhove L, Tack J. Serotonin transporter gene polymorphisms in irritable bowel syndrome. Neurogastroenterol Motil 2006; 18(11): 957-959.

9. Heiser P, Dempfle A, Friedel S, Konrad K, Hinney A, Kiefl H, Walitza S, Bettecken T, Saar K, Linder M, Warnke A, Herpertz-Dahlmann B, Schäfer H, Remschmidt H, Hebebrand J. Family-based association study of serotonergic candidate genes and attention-deficit/hyperactivity disorder in a German sample. J Neural Transm 2007; 114 (4): 513-521.

10. Pata C, Erdal E, Yazc K, Camdeviren H, Ozkaya M, Ulu O. Association of the -1438 G/A and 102 T/C polymorphism of the 5-Ht2A receptor gene with irritable bowel syndrome 5-Ht2A gene polymorphism in irritable bowel syndrome. J Clin Gastroenterol 2004; 38(7): 561-566.

11. Saito YA, Cremonini F, Talley NJ. Association of the 1438G/A and 102T/C polymorphism of the 5-HT2A receptor gene with irritable bowel syndrome 5-HT2A gene polymorphism in irritable bowel syndrome. J Clin Gastroenterol 2005; 39(9): 835.

12. Anttila S, Kampman O, Illi A, Rontu R, Lehtimäki T, Leinonen E. Association between 5-HT2A, TPH1 and GNB3 genotypes and response to typical neuroleptics: a serotonergic approach. BMC Psychiatry 2007; 237: 22.

13. Krzywkowski K, Jensen AA, Connolly CN, Bräuner-Osborne H. Naturally occurring variations in the human 5-HT3A gene profoundly

impact 5-HT3 receptor function and expression. Pharmacogenet Genomics 2007; 17 (4): 255-266.

14. Kim HJ, Camilleri M, Carlson PJ, Cremonini F, Ferber I, Stephens D, McKinzie S, Zinsmeister AR, Urrutia R. Association of distinct alpha(2) adrenoceptor and serotonin transporter polymorphisms with constipationd somatic symptoms in functional gastrointestinal disorders. Gut 2004; 53(6): 829-837.

15. Saito YA, Locke GR, Zimmerman JM, Holtmann G, Slusser JP, de Andrade M, Petersen GM, Talley NJ. A genetic association study of 5-HTT LPR and GNbeta3 C825T polymorphisms with irritable bowel syndrome. Neurogastroenterol Motil 2007; 19 (6): 465-470.

16. Andresen V, Camilleri M, Kim HJ, Stephens DA, Carlson PJ, Talley NJ, Saito YA, Urrutia R, Zinsmeister AR. Is there an association between GNbeta3-C825T genotype and lower functional gastrointestinal disorders? Gastroenterology 2006; 130(7): 1985-1994.

17. Cremonini F, Camilleri M, McKinzie S, Carlson P, Camilleri CE, Burton D, Thomforde G, Urrutia R, Zinsmeister AR. Effect of CCK-1 antagonist, dexloxiglumide, in female patients with irritable bowel syndrome: a pharmacodynamic and pharmacogenomic study. Am J Gastroenterol 2005; 100 (3): 652-663.

18. Schulte CM, Goebell H, Roher HD, Schulte KM. C-509T polymorphism in the TGFB1 gene promoter: impact on Crohn's disease susceptibility and clinical course? Immunogenetics 2001; 53(2): 178-182.

19. Hume GE, Fowler EV, Lincoln D, Eri R, Templeton D, Florin TH, Cavanaugh JA, Radford-Smith GL. Angiotensinogen and transforming

growth factor beta1: novel genes in the pathogenesis of Crohn's disease. J Med Genet 2006; 43(10): e51.

20. Tamizifar B, Bagheri-Lankarani K, Naeemi S, Rismankar-Zadeh M, Taghavi A,Ghaderi A. Polymorphism of the promoter region of C-509T of transforming growth factor-beta1 gene and ulcerative colitis. Arch Iran Med 2007; 10(2): 171-175.

21. Scheuch K, Lautenschlager M, Grohmann M, Stahlberg S, Kirchheiner J, Zill P, Heinz A, Walther DJ, Priller J. Characterization of a functional Promoter Polymorphism of the Human Tryptophan Hydroxylase 2 Gene in Serotonergic Raphe Neurons. Biol Psychiatry 2007; 62: 1288-1294.

22. Walitza S, Renner TJ, Dempfle A, Konrad K, Wewetzer Ch, Halbach A, Herpertz-Dahlmann B, Remschmidt H, Smidt J, Linder M, Flierl L, Knölker U, Friedel S, Schäfer H, Gross C, Hebebrand J, Warnke A, Lesch KP. Transmission disequilibrium of polymorphic variants in the tryptophan hydroxylase-2 gene in attention-deficit/hyperactivity disorder. Mol Psychiatry 2005; 10(12): 1126-1132.

23. Van Den Bogaert A, Sleegers K, De Zutter S, Heyrman L, Norrback KF, Adolfsson R, Van Broeckhoven C, Del-Favero J. Association of brain-specific tryptophan hydroxylase, TPH2, with unipolar and bipolar disorder in a Northern Swedish, isolated population. Arch Gen Psychiatry 2006; 63(10): 1103-1110.

24. Chen GL, Vallender EJ, Miller GM. Functional characterization of the human TPH2 5' regulatory region: untranslated region and polymorphisms modulate gene expression in vitro. Hum Genet 2008; 122(6): 645-657.

25. Cantor MJ, Nickerson P, Bernstein CN. The role of cytokine gene polymorphisms in determining disease susceptibility and phenotype in inflammatory bowel disease. Am J Gastroenterol 2005; 100: 1134-1142.

26. Balding J, Livingstone WJ, Conroy J, Mynett-Johnson L, Weir DG, Mahmud N, Smith OP. Inflammatory bowel disease: the role of inflammatory cytokine gene polymorphisms. Mediators Inflamm 2004; 13(3): 181-187.

27. Tagore A, Gonsalkorale WM, Pravica V, Hajeer AH, McMahon R, Whorwell PJ, Sinnott PJ, Hutchinson IV. Interleukin-10 (IL-10) genotypes in inflammatory bowel disease. Tissue Antigens 1999; 54(4): 386-390.

28. Aithal GP, Craggs A, Day CP, Welfare M, Daly AK, Mansfield JC, Hudson M. Role of polymorphisms in the interleukin-10 gene in determining disease susceptibility and phenotype in inflamatory bowel disease. Dig Dis Sci 2001; 46(7): 1520-1525.

29. Opdal SH.IL-10 gene polymorphisms in infectious disease and SIDS. FEMS Immunol Med Microbiol 2004; 42(1): 48-52.

30. Castro-Santos P, Suarez A, Lopez-Rivas L, Mozo L, Gutierrez C. TNFalpha and IL-10 gene polymorphisms in inflammatory bowel disease. Association of -1082 AA low producer IL-10 genotype with steroid dependency. Am J Gastroenterol 2006; 101(5): 1039-1047.

31. van der Veek PP, van den Berg M, de Kroon YE, Verspaget HW, Masclee AA. Role of tumor necrosis factor-alpha and interleukin-10 gene polymorphisms in irritable bowel syndrome. Am J Gastroenterol 2005; 100(11): 2510-2516.

32. Gonsalkorale WM, Perrey C, Pravica V, Whorwell PJ, Hutchinson IV. Interleukin 10 genotypes in irritable bowel syndrome: evidence for an inflammatory component? Gut 2003; 52(1): 91-93.

33. Collins SM. Dysregulation of peripheral cytokine production in irritable bowel syndrome. Am J Gastroenterol 2005; 100(11): 2517-2518.

34. Peltekova VD, Wintle RF, Rubin LA, Amos CI, Huang Q, Gu X, Newman B, Van Oene M,Cescon D, Greenberg G, Griffiths AM, St George-Hyslop PH, Siminovitch KA. Functional variants of OCTN cation transporter genes are associated with Crohn disease. Nat Genet 2004; 36(5): 471-475.

35. Kiemeney LA, van Houwelingen KP, Bogaerts M, Witjes JA, Swinkels DW, den Heijer M, Franke B, Schalken JA, Gerald W. Verhaegh. Polymorphisms in the E-cadherin (CDH1) gene promoter and the risk of bladder cancer. Eur J Cancer 2006; 42 (18): 3219-3227.

36. Li HC, Albert JM, Shinohara ET, Cai Q, Freyer A, Cai H, Cao C, Wang Z, Kataoka N, Teng M, Zheng W, Lu B. E-cadherin promoter polymorphisms are not associated with the aggressiveness of prostate cancer in Caucasian patients. Urol Oncol 2006; 24(6): 496-502.

37. Lindström S, Wiklund F, Jonsson BA, Adami HO, Bälter K, Brookes AJ, Xu J, Zheng SL, Isaacs WB, Adolfsson J, Grönberg H. Comprehensive genetic evaluation of common E-cadherin sequence variants and prostate cancer risk: strong confirmation of functional promoter SNP. Hum Genet 2005; 118 (3-4): 339-347.

38. Braun C, Zahn R, Martin K, Albert E, Folwaczny C. Polymorphisms of the ICAM-1 gene are associated with inflammatory bowel disease, regardless of the p-ANCA status. Clin Immunol 2001; 101(3): 357-360.

39. Matsuzawa J, Sugimura K, Matsuda Y, Takazoe M, Ishizuka K, Mochizuki T, Seki SS,Yoneyama O, Bannnai H, Suzuki K, Honma T, Asakura H. Association between K469E allele of intercellular adhesion molecule 1 gene and inflammatory bowel disease in a Japanese population. Gut 2003; 52(1): 75-78.

40. Low JH, Williams FA, Yang X, Cullen S, Colley J, Ling KL, Armuzzi A, Ahmad T,Neville MJ, Dechairo BM, Walton R, Lench NJ, Jewell DP. Inflammatory bowel disease is linked to 19p13 and associated with ICAM-1. Inflamm Bowel Dis 2004; 10(3): 173-181.

41. Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA,Rosenstiel P, Albrecht M, Croucher PJ, Seegert D, Nikolaus S, Hampe J, Lengauer T, Pierrou S, Foelsch UR, Mathew CG, Lagerstrom-Fermer M, Schreiber S. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nat Genet 2004; 36(5): 476-480.

42. van Bodegraven AA, Curley CR, Hunt KA, Monsuur AJ, Linskens RK, Onnie CM,Crusius JB, Annese V, Latiano A, Silverberg MS, Bitton A, Fisher SA, Steinhart AH, Forbes A, Sanderson J, Prescott NJ, Strachan DP, Playford RJ, Mathew CG, Wijmenga C, Daly MJ, Rioux JD, van Heel DA. Genetic variation in myosin IXB is associated with ulcerative colitis. Gastroenterology 2006; 131(6): 1768-1774.

43. Brant SR, Panhuysen CI, Nicolae D, Reddy DM, Bonen DK, Karaliukas R, Zhang L, Swanson E, Datta LW, Moran T, Ravenhill G, Duerr RH, Achkar JP, Karban AS, Cho JH. MDR1 Ala893 polymorphism is

associated with inflammatory bowel disease. Am J Hum Genet 2003; 73(6): 1282-1292.

44. Ho GT, Soranzo N, Nimmo ER, Tenesa A, Goldstein DB, Satsangi J. ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis:discrimination of critical variants using a gene-wide haplotype tagging approach. Hum Mol Genet 2006; 15(5): 797-805.

45. Annese V, Valvano MR, Palmieri O, Latiano A, Bossa F, Andriulli A. Multidrug resistance 1 gene in inflammatory bowel disease: a meta-analysis. World J Gastroenterol 2006; 12(23): 3636-3644.

46. Dring MM, Goulding CA, Trimble VI, Keegan D, Ryan AW, Brophy KM, Smyth CM,Keeling PW, O'Donoghue D, O'Sullivan M, O'Morain C, Mahmud N, Wikstrom AC,Kelleher D, McManus R. The pregnane X receptor locus is associated with susceptibility to inflammatory bowel disease. Gastroenterology 2006; 130(2): 341-348.

47. Mohamed JA, DuPont HL, Jiang ZD, Belkind-Gerson J, Figueroa JF, Armitige LY, Tsai A, Nair P, Martinez-Sandoval FJ, Guo DC, Hayes P, Okhuysen PC. A novel single-nucleotide polymorphism in the lactoferrin gene is associated with susceptibility to diarrhea in North American travelers to Mexico. Clin Infect Dis 2007; 44(7): 945-952.

48. Klein W, Tromm A, Folwaczny C, Hagedorn M, Duerig N, Epplen J, Schmiegel W,Griga T. A polymorphism of the bactericidal/permeability increasing protein (BPI) gene is associated with Crohn's disease. J Clin Gastroenterol 2005; 39(4): 282-283.

49. Monajemi H, Meenan J, Lamping R, Obradov DO, Radema SA, Trown PW, Tytgat GN, Van Deventer SJ. Inflammatory bowel disease is

associated with increased mucosal levels of bactericidal/permeability-increasing protein. Gastroenterology 1996; 110(3): 733-739.

50. Hubacek JA, Stuber F, Frohlich D, Book M, Wetegrove S, Ritter M, Rothe G,Schmitz G. Gene variants of the bactericidal/permeability increasing protein and lipopolysaccharide binding protein in sepsis patients: gender-specific genetic predisposition to sepsis. Crit Care Med 2001; 29(3): 557-561.

51. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S,Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R,Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001; 411 (6837): 599-603.

52. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001; 411(6837): 603-606.

53. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A, Demarche B, Gut I, Heath S, Foglio M, Liang L,Laukens D, Mni M, Zelenika D, Gossum AV, Rutgeerts P, Belaiche J, Lathrop M,Georges M. 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): e58.

54. McGovern DP, Hysi P, Ahmad T, van Heel DA, Moffatt MF, Carey A, Cookson WO, Jewell DP. Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. Hum Mol Genet 2005; 14(10): 1245-1250.

55. McGovern DP, Butler H, Ahmad T, Paolucci M, van Heel DA, Negoro K, Hysi P,Ragoussis J, Travis SP, Cardon LR, Jewell DP. TUCAN (CARD8) genetic variants and inflammatory bowel disease. Gastroenterology 2006; 131(4): 1190-1196.

56. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, Frees K, Watt JL, Schwartz DA. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. Nat Genet 2000; 25(2): 187-191.

57. Franchimont D, Vermeire S, El Housni H, Pierik M, Van Steen K, Gustot T, Quertinmont E, Abramowicz M, Van Gossum A, Deviere J, Rutgeerts P. 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): 987-992.

58. Marsik C, Jilma B, Joukhadar C, Mannhalter C, Wagner O, Endler G. The Toll-like receptor 4 Asp299Gly and Thr399Ile polymorphisms influence the late inflammatory response in human endotoxemia. Clin Chem 2005; 51(11): 2178-2180.

59. Candore G, Aquino A, Balistreri CR, Bulati M, Di Carlo D, Grimaldi MP, Listi F,Orlando V, Vasto S, Caruso M, Colonna-Romano G, Lio D,

Caruso C. Inflammation, longevity, and cardiovascular diseases: role of polymorphisms of TLR4. Ann N Y Acad Sci 2006; 1067: 282-287.

60. Hawn TR, Verbon A, Lettinga KD, Zhao LP, Li SS, Laws RJ, Skerrett SJ, Beutler B,Schroeder L, Nachman A, Ozinsky A, Smith KD, Aderem A. TLR5 signaling and is associated with susceptibility to legionnaires' disease. J Exp Med 2003; 198(10): 1563-1572.

61. Gewirtz AT, Vijay-Kumar M, Brant SR, Duerr RH, Nicolae DL, Cho JH. Dominant-negative TLR5 polymorphism reduces adaptive immune response to flagellin and negatively associates with Crohn's disease. Am J Physiol Gastrointest Liver Physiol 2006; 290(6): G1157-1163.

62. Torok HP, Glas J, Tonenchi L, Bruennler G, Folwaczny M, Folwaczny C. Crohn's disease is associated with a toll-like receptor-9 polymorphism. Gastroenterology 2004; 127(1): 365-366.

63. Lazarus R, Klimecki WT, Raby BA, Vercelli D, Palmer LJ, Kwiatkowski DJ, Silverman EK, Martinez F, Weiss ST. Single-nucleotide polymorphisms in the Toll-like receptor 9 gene (TLR9): frequencies, pairwise linkage disequilibrium, and haplotypes in three U.S. ethnic groups and exploratory case-control disease association studies. Genomics 2003; 81(1): 85-91.

64. Obana N, Takahashi S, Kinouchi Y, Negoro K, Takagi S, Hiwatashi N, Shimosegawa T. Ulcerative colitis is associated with a promoter polymorphism of lipopolysaccharide receptor gene, CD14. Scand J Gastroenterol 2002; 37(6): 699-704.

65. Klein W, Tromm A, Griga T, Fricke H, Folwaczny C, Hocke M, Eitner K, Marx M,Duerig N, Epplen JT. A polymorphism in the CD14 gene is

associated with Crohn disease. Scand J Gastroenterol 2002; 37(2): 189-191.

66. Arnott ID, Nimmo ER, Drummond HE, Fennell J, Smith BR, MacKinlay E, Morecroft J,Anderson N, Kelleher D, O'Sullivan M, McManus R, Satsangi J. NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? Genes Immun 2004; 5(5): 417-425.

67. Temple SE, Cheong KY, Almeida CM, Price P, Waterer GW. Polymorphisms in lymphotoxin alpha and CD14 genes influence TNFalpha production induced by Gram-positive and Gram-negative bacteria. Genes Immun 2003; 4(4): 283-288.

68. Karban AS, Okazaki T, Panhuysen CI, Gallegos T, Potter JJ, Bailey-Wilson JE,Silverberg MS, Duerr RH, Cho JH, Gregersen PK, Wu Y, Achkar JP, Dassopoulos T,Mezey E, Bayless TM, Nouvet FJ, Brant SR. Functional annotation of a novel NFKB1 promoter polymorphism that increases risk for ulcerative colitis. Hum Mol Genet 2004; 13(1): 35-45.

69. Borm ME, van Bodegraven AA, Mulder CJ, Kraal G, Bouma G. A NFKB1 promoter polymorphism is involved in susceptibility to ulcerative colitis. Int J Immunogenet 2005; 32(6): 401-405.

70. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G,De La Vega FM, Briggs J, Gunther S, Prescott NJ, Onnie CM, Hasler R, Sipos B,Folsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 2007; 39(2): 207-211.

71. The 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): 661-678.

72. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 2007; 39(5): 596-604.

73. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhart AH, Targan SR, Xavier RJ; NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008; 40(8):955-962.

74. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R,Burke DA, Todhunter CE, Ahmad T, Onnie CM, McArdle W, Strachan D, Bethel G, Bryan C, Lewis CM,

Deloukas P, Forbes A, Sanderson J, Jewell DP, Satsangi J, Mansfield JC; Wellcome Trust Case Control Consortium, Cardon L, Mathew CG. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 2007; 39(7): 830-832.

75. Negoro K, Kinouchi Y, Hiwatashi N, Takahashi S, Takagi S, Satoh J, Shimosegawa T, Toyota T. Crohn's disease is associated with novel polymorphisms in the 5'-flanking region of the tumor necrosis factor gene. Gastroenterology 1999; 117(5): 1062-1068.

76. Skoog T, van't Hooft FM, Kallin B, Jovinge S, Boquist S, Nilsson J, Eriksson P,Hamsten A. A common functional polymorphism (C-->A substitution at position -863) in the promoter region of the tumour necrosis factor-alpha (TNF-alpha) gene associated with reduced circulating levels of TNF-alpha. Hum Mol Genet 1999; 8(8): 1443-1449.

77. Sashio H, Tamura K, Ito R, Yamamoto Y, Bamba H, Kosaka T, Fukui S, Sawada K,Fukuda Y, Tamura K, Satomi M, Shimoyama T, Furuyama J. Polymorphisms of the TNF gene and the TNF receptor superfamily member 1B gene are associated with susceptibility to ulcerative colitis and Crohn's disease,respectively. Immunogenetics 2002; 53(12): 1020-1027.

78. van Heel DA, Udalova IA, De Silva AP, McGovern DP, Kinouchi Y, Hull J, Lench NJ,Cardon LR, Carey AH, Jewell DP, Kwiatkowski D. Inflammatory bowel disease is associated with a TNF polymorphism that affects an interaction between the OCT1 and NF(-kappa)B transcription factors. Hum Mol Genet 2002; 11(11): 1281-1289.

79. Bioque G, Crusius JB, Koutroubakis I, Bouma G, Kostense PJ, Meuwissen SG, Pena AS. Allelic polymorphism in IL-1 beta and IL-1 receptor antagonist (IL-1Ra) genes in inflammatory bowel disease. Clin Exp Immunol 1995; 102(2): 379-383.

80. Heresbach D, Alizadeh M, Dabadie A, Le Berre N, Colombel JF, Yaouanq J, Bretagne JF, Semana G. Significance of interleukin-1beta and interleukin-1 receptor antagonist genetic polymorphism in inflammatory bowel diseases. Am J Gastroenterol 1997; 92(7): 1164-1169.

81. Cox A, Camp NJ, Nicklin MJ, di Giovine FS, Duff GW. An analysis of linkage disequilibrium in the interleukin-1 gene cluster, using a novel grouping method for multiallelic markers. Am J Hum Genet 1998; 62(5): 1180-1188.

82. Nemetz A, Nosti-Escanilla MP, Molnar T, Kope A, Kovacs A, Feher J, Tulassay Z, Nagy F, Garcia-Gonzalez MA, Pena AS. IL1B gene polymorphisms influence the course and severity of inflammatory bowel disease. Immunogenetics 1999; 49(6): 527-531.

83. El-Omar EM, Carrington M, Chow WH, McColl KE, Bream JH, Young HA, Herrera J,Lissowska J, Yuan CC, Rothman N, Lanyon G, Martin M, Fraumeni JF Jr, Rabkin CS. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature 2000; 404(6776): 398-402.

84. Mwantembe O, Gaillard MC, Barkhuizen M, Pillay V, Berry SD, Dewar JB, Song E. Ethnic differences in allelic associations of the interleukin-1 gene cluster in South African patients with inflammatory bowel disease (IBD) and in control individuals. Immunogenetics 2001; 52(3-4): 249-254.

85. Nohara H, Saito Y, Higaki S, Okayama N, Hamanaka Y, Okita K, Hinoda Y. Polymorphisms of the IL-1beta and IL-1beta-inducible genes in ulcerative colitis. J Gastroenterol 2002; 37 (Suppl 14): 107-110.

86. van der Paardt M, Crusius JB, Garcia-Gonzalez MA, Baudoin P, Kostense PJ,Alizadeh BZ, Dijkmans BA, Pena AS, van der Horst-Bruinsma IE. Interleukin-1beta and interleukin-1 receptor antagonist gene polymorphisms in ankylosing spondylitis. Rheumatology (Oxford). 2002; 41(12): 1419-1423.

87. Liou JM, Lin JT, Wang HP, Huang SP, Lee YC, Chiu HM, Shun CT, Wu MS. IL-1B-511 C-->T polymorphism is associated with increased host susceptibility to Helicobacter pylori infection in Chinese. Helicobacter 2007; 12(2): 142-149.

88. Klein W, Tromm A, Griga T, Fricke H, Folwaczny C, Hocke M, Eitner K, Marx M,Duerig N, Epplen JT. Interleukin-4 and interleukin-4 receptor gene polymorphisms in inflammatory bowel diseases. Genes Immun 2001; 2(5): 287-289.

89. Aithal GP, Day CP, Leathart J, Daly AK, Hudson M. Association of single nucleotide polymorphisms in the interleukin-4 gene and interleukin-4 receptor gene with Crohn's disease in a British population. Genes Immun 2001; 2(1): 44-47.

90. Landi S, Bottari F, Gemignani F, Gioia-Patricola L, Guino E, Osorio A, de Oca J, Capella G, Canzian F, Moreno V; Bellvitge Colorectal Cancer Study Group. Interleukin-4 and interleukin-4 receptor polymorphisms and colorectal cancer risk. Eur J Cancer 2007; 43 (4): 762-768.

91. López KI, Martinez SE, Moguel MC, Romero LT, Figueroa CS, Pacheco GV, Ibarra B, Corona JS. Genetic diversity of the IL-4, IL-4 receptor and IL-13 loci in mestizos in the general population and in patients with asthma from three subpopulations in Mexico. Int J Immunogenet 2007; 34 (1): 27-33.

92. Wenzel SE, Balzar S, Ampleford E, Hawkins GA, Busse WW, Calhoun WJ, Castro M,Chung KF, Erzurum S, Gaston B, Israel E, Teague WG, Curran- Everett D, Meyers DA, Bleecker ER. IL4Ralpha Mutations are associated with asthma exacerbations and mast cell/IgE expression. Am J Respir Crit Care Med 2007; 175 (6): 570-576.

93. Hytönen AM, Löwhagen O, Arvidsson M, Balder B, Björk AL, Lindgren S, Hahn-Zoric M, Hanson LA, Padyukov L. Haplotypes of the interleukin-4 receptor alpha chain gene associate with susceptibility to and severity of atopic asthma. Clin Exp Allergy 2004; 34(10): 1570-1575.

94. Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, Woo P. The effect of novel polymorphisms in the interleukin-6 (*IL6*) gene on *IL6* transcription and plasma *IL6* levels, and an association with systemic-onset juvenile chronic arthritis. J Clin Invest 1998; 102(7): 1369-1376.

95. Rafiq S, Frayling TM, Murray A, Hurst A, Stevens K, Weedon MN, Henley W, Ferrucci L, Bandinelli S, Corsi AM, Guralnik JM, Melzer D. A common variant of the interleukin 6 receptor (IL-6r) gene increases IL-6r and IL-6 levels, without other inflammatory effects. Genes Immun 2007; 8(7): 552-559

96. Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, Caillier SJ, Ban M, Goris A, Barcellos LF, Lincoln R, McCauley JL, Sawcer SJ, Compston DA, Dubois B, Hauser SL, Garcia-Blanco MA, Pericak-Vance MA, Haines JL; Multiple Sclerosis Genetics Group. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet 2007; 39(9): 1083-1091.

97. Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallström E, Khademi M, Oturai A, Ryder LP, Saarela J, Harbo HF, Celius EG, Salter H, Olsson T, Hillert J. Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. Nat Genet 2007; 39(9): 1108-1113.

98. Jiang ZD, Okhuysen PC, Guo DC, He R, King TM, DuPont HL, Milewicz DM. Genetic susceptibility to enteroaggregative Escherichia coli diarrhea: polymorphism in the interleukin-8 promotor region. J Infect Dis 2003; 188(4): 506-511.

99. Jiang ZD, DuPont HL, Garey K, Price M, Graham G, Okhuysen P, Dao-Tran T, LaRocco M. A common polymorphism in the interleukin 8 gene promoter is associated with Clostridium difficile diarrhea. Am J Gastroenterol 2006; 101(5): 1112-1116.

100. Gyulai Z, Klausz G, Tiszai A, Lenart Z, Kasa IT, Lonovics J, Mandi Y. Genetic polymorphism of interleukin-8 (IL-8) is associated with Helicobacter pylori-induced duodenal ulcer. Eur Cytokine Netw 2004; 15(4): 353-358.

101. Zwiers A, Seegers D, Heijmans R, Koch A, Hampe J, Nikolaus S, Pena AS, Schreiber S, Bouma G. Definition of polymorphisms and haplotypes in the interleukin-12B gene:association with IL-12

production but not with Crohn's disease. Genes Immun 2004; 5(8): 675-677.

102. Hirota T, Suzuki Y, Hasegawa K, Obara K, Matsuda A, Akahoshi M, Nakashima K,Cheng L, Takahashi N, Shimizu M, Doi S, Fujita K, Enomoto T, Ebisawa M, Yoshihara S, Nakamura Y, Kishi F, Shirakawa T, Tamari M. Functional haplotypes of IL-12B are associated with childhood atopic asthma. J Allergy Clin Immunol 2005; 116(4): 789-795.

103. Howard TD, Whittaker PA, Zaiman AL, Koppelman GH, Xu J, Hanley MT, Meyers DA, Postma DS, Bleecker ER. Identification and association of polymorphisms in the interleukin-13 gene with asthma and atopy in a Dutch population. Am J Respir Cell Mol Biol 2001; 25(3): 377-384.

104. Howard TD, Koppelman GH, Xu J, Zheng SL, Postma DS, Meyers DA, Bleecker ER. Gene-gene interaction in asthma: IL4RA and IL13 in a Dutch population with asthma. Am J Hum Genet 2002; 70(1): 230-236.

105. He JQ, Chan-Yeung M, Becker AB, Dimich-Ward H, Ferguson AC, Manfreda J, Watson WT, Sandford AJ. Genetic variants of the IL13 and IL4 genes and atopic diseases in at-risk children. Genes Immun 2003; 4(5): 385-389.

106. Heinzmann A, Mao XQ, Akaiwa M, Kreomer RT, Gao PS, Ohshima K, Umeshita R, Abe Y, Braun S, Yamashita T, Roberts MH, Sugimoto R, Arima K, Arinobu Y, Yu B, Kruse S, Enomoto T, Dake Y, Kawai M, Shimazu S, Sasaki S, Adra CN, Kitaichi M, Inoue H, Yamauchi K, Tomichi N, Kurimoto F, Hamasaki N, Hopkin JM, Izuhara K,

Shirakawa T, Deichmann KA. Genetic variants of IL-13 signalling and human asthma and atopy. Hum Mol Genet 2000; 9(4): 549-559.

107. Vladich FD, Brazille SM, Stern D, Peck ML, Ghittoni R, Vercelli D. IL-13 R130Q, a common variant associated with allergy and asthma, enhances effector mechanisms essential for human allergic inflammation. J Clin Invest 2005; 115(3): 747-754.

108. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH,Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S,Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI,Nicolae DL, Cho JH. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006; 314(5804): 1461-1463.

109. De Jager PL, Richardson A, Vyse TJ, Rioux JD. Genetic variation in toll-like receptor 9 and susceptibility to systemic lupus erythematosus. Arthritis Rheum 2006; 54(4): 1279-1282.

110. Lazarus R, Klimecki WT, Raby BA, Vercelli D, Palmer LJ, Kwiatkowski DJ, Silverman EK, Martinez F, Weiss ST. Single-nucleotide polymorphisms in the Toll-like receptor 9 gene (TLR9): frequencies, pairwise linkage disequilibrium, and haplotypes in three U.S. ethnic groups and exploratory case-control disease association studies. Genomics 2003; 81(1): 85-91.

111. Berghöfer B, Frommer T, König IR, Ziegler A, Chakraborty T, Bein G, Hackstein H. Common human Toll-like receptor 9 polymorphisms and haplotypes: association with atopy and functional relevance. Clin Exp Allergy 2005; 35(9): 1147-1154.

112. Novak N, Yu CF, Bussmann C, Maintz L, Peng WM, Hart J, Hagemann T, Diaz-Lacava A, Baurecht HJ, Klopp N, Wagenpfeil S, Behrendt H, Bieber T, Ring J, Illig T, Weidinger S. Putative association of a TLR9 promoter polymorphism with atopic eczema. Allergy 2007; 62(7): 766-772.

113. Kikuchi K, Lian ZX, Kimura Y, Selmi C, Yang GX, Gordon SC, Invernizzi P, Podda M, Coppel RL, Ansari AA, Ikehara S, Miyakawa H, Gershwin ME. Genetic polymorphisms of toll-like receptor 9 influence the immune response to CpG and contribute to hyper-IgM in primary biliary cirrhosis. J Autoimmun 2005; 24(4): 347-352.

114. Hong J, Leung E, Fraser AG, Merriman TR, Vishnu P, Krissansen GW. TLR2, TLR4 and TLR9 polymorphisms and Crohn's disease in a New Zealand Caucasian cohort. J Gastroenterol Hepatol 2007; 22(11): 1760-1766.

115. Carvalho A, Santos M, Maciel P, Rodrigues F. T-1237C polymorphism of TLR9 gene is not associated with multiple sclerosis in the Portuguese population. Mult Scler 2008; 14(4): 550-552.

116. Fife MS, Ogilvie EM, Kelberman D, Samuel J, Gutierrez A, Humphries SE, Woo P. Novel IL-6 haplotypes and disease association. Genes Immun 2005; 6(4): 367-370.

117. Qi L, van Dam RM, Meigs JB, Manson JE, Hunter D, Hu FB. Genetic variation in IL6 gene and type 2 diabetes: tagging-SNP haplotype analysis in large-scale case-control study and meta-analysis. Hum Mol Genet 2006; 15(11): 1914-1920.

118. Liu Y, Berthier-Schaad Y, Fallin MD, Fink NE, Tracy RP, Klag MJ, Smith MW, Coresh J. IL-6 haplotypes, inflammation, and risk for cardiovascular disease in a multiethnic dialysis cohort. J Am Soc Nephrol 2006; 17(3): 863-870.

119. Müller-Steinhardt M, Fricke L, Müller B, Ebel B, Kirchner H, Härtel C. Cooperative influence of the interleukin-6 promoter polymorphisms -597,-572 and -174 on long-term kidney allograft survival. Am J Transplant 2004; 4(3): 402-406.

120. Müller-Steinhardt M, Ebel B, Härtel C. The impact of interleukin-6 promoter -597/-572/-174genotype on interleukin-6 production after lipopolysaccharide stimulation. Clin Exp Immunol 2007; 147(2): 339-345.

121. Taudorf S, Krabbe KS, Berg RM, Møller K, Pedersen BK, Bruunsgaard. Common studied polymorphisms do not affect plasma cytokine levels upon endotoxin exposure in humans. Clin Exp Immunol 2008; 152(1): 147-152.

122. Shin Y, Kim IJ, Kang HC, Park JH, Park HR, Park HW, Park MA, Lee JS, Yoon KA, Ku JL, Park JG. The E-cadherin -347G->GA promoter polymorphism and its effect on transcriptional regulation. Carcinogenesis 2004; 25(6): 895-899.

123. Nakamura A, T, Kaneko K, Shibata M, Matsumura T, Nagai M, Makino R, Mitamura K. Characterization of DNA polymorphisms in the E-cadherin gene (CDH1) promoter region. Mutat Res 2002; 502: 19-24.

124. Li LC, Chui RM, Sasaki M, Nakajima K, Perinchery G, Au HC, Nojima D, Carroll P,Dahiya R. A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities. Cancer Res 2000; 60(4): 873-876.

125. Cattaneo F, Venesio T, Molatore S, Russo A, Fiocca R, Frattini M, Scovassi AI, Ottini L, Bertario L, Ranzani GN. Functional analysis and case-control study of -160C/A polymorphism in the E-cadherin gene promoter: association with cancer risk. Anticancer Res 2006 ; 26(6B): 4627-4632.

126. Bonilla C, Mason T, Long L, Ahaghotu C, Chen W, Zhao A, Coulibaly A, Bennett F, Aiken W, Tullock T, Coard K, Freeman V, Kittles RA. E-cadherin polymorphisms and haplotypes influence risk for prostate cancer. Prostate 2006; 66(5): 546-556.

127. Lindström S, Wiklund F, Jonsson BA, Adami HO, Bälter K, Brookes AJ, Xu J, Zheng SL, Isaacs WB, Adolfsson J, Grönberg H. Comprehensive genetic evaluation of common E-cadherin sequence variants and prostate cancer risk: strong confirmation of functional promoter SNP. Hum Genet 2005; 118: 339-347.

# 6 SUMMARY & PERSPECTIVES

## 6.1 General discussion of the chapters

This thesis describes studies aimed at further understanding the role of innate immunity genetic risk factors involved in the origin and development of IBD and PI-IBS. Using a candidate gene approach, we have successfully uncovered a new susceptibility gene, *NLRP3*, contributing to CD pathogenesis by dissecting two gene regions involved in the inflammasome pathway, and have established the first basis of the genetic risk factors (namely *TLR9*, *CDH1*, and *IL6*) involved in PI-IBS susceptibility.

### 6.1.1 CD immunopathogenesis

#### 6.1.1.1 Association results in perspectives

The *NLRP3* association results presented in this thesis support the recent shift in the immunopathogenesis paradigm of CD, which is based on the concept that mediators of the innate immunity and abnormalities in the handling of intracellular bacteria are emerging as a key theme in CD pathogenesis.

CD is known to be a polygenic and heterogeneous disease with a strong environmental influence. It is a condition that is manifested with ulceration and chronic inflammation of the GI tract (Podolsky 2002). In the established tissue lesions observed in patients, there is an infiltration of inflammatory cells focally in the bowel wall, leading to the production of pro-inflammatory cytokines and propagation of the immune response (Podolsky 2002). These observations have

led researchers to initially hypothesise that CD was the result of an overly aggressive inflammatory response and uncontrolled production of pro-inflammatory mediators. This hypothesis is further supported by the successful and beneficial use of anti-TNF$\alpha$ biologic therapies and new selective adhesion molecule inhibitors in CD (Derkx et al. 1993; van Dullemen et al. 1995; Targan et al. 1997; Kozuch et al. 2008), targeting the excessive immune activation.

However, the publications of several reports supporting a "loss-of-function" phenotype associated with the NOD2 mutations, together with the recently published GWA studies highlighting the role of the autophagy pathway (*IRGM* and *ATG16L1*) in the CD pathogenesis, have revolutionized the CD immunopathogenesis paradigm by suggesting that the pro-inflammatory response observed in CD patients might be secondary to the primary inability to handle the bacterial burden from the intestinal lumen. It is now suggested that an impaired, rather than an overly aggressive inflammatory response, by a defective intestinal innate immune system may underlie the initial phase of IBD, most probably resulting from an inability to handle occasional invaders from the intestinal lumen. A primary immunodeficiency defect may result to failure of producing inflammatory mediators, leading to insufficient recruitment of immune effector cells like neutrophils, resulting in inadequate removal of bacteria and antigens. This impairment of acute inflammatory response can be compensated in some circumstances by signaling through PRMs, such as NOD2 and NALP3. If these PRMs do not recognize properly and initiate the mechanisms to clear the threat, the foreign material in the bowel wall may be taken up by macrophages, eliciting a granulomatous reaction and the local and systemic sequelae that is characteristic

of CD. Activation of the acquired immune system would thus be a subsequent phenomenon, mostly associated with the perpetuation of inflammation, being the proximate driver of tissue damage observed in patients (Xavier et al. 2007).

Following this shift in the pathogenesis paradigm, new treatment strategies should aim at targeting the microbial flora and at re-setting the GI mucosal immune system towards this flora using immuno-modulating drugs (with both permissive and negative actions) more than the immuno-suppressive drugs (only negative actions) that are currently being used, such as antimetabolites and biological therapies.

Probiotics are live and safe microbes that may beneficially help in restoring the intestinal microbial flora (Fedorak et al. 2008). Two meta-analyses failed to show the beneficial use of probiotics in the maintenance of remission in CD patients (Rolfe et al. 2006; Rahimi et al. 2008). Overall, studies evaluating the efficacy of probiotics as induction and maintenance therapies have been quite disappointing in CD (Guslandi et al. 2000; Prantera et al. 2002; Bousvaros et al. 2005; Marteau et al. 2006; Van Gossum et al. 2007). However, as most of these probiotic studies were small opened label or insufficiently powered randomized controlled trials, no definitive conclusions can be drawn. The inconsistent results observed from the clinical trials on probiotics might be explained by the heterogeneity and variety of probiotic preparations used, study designs, the medication regimen, together with the poor selection of patient study populations (Fedorak et al. 2008). A better characterization of the gut microbial flora and its genome, together with a better understanding of its dynamic interactions with the human host and its genome, would be needed before pursuing further evaluation

of the efficacy of specific probiotics in larger randomized clinical trials. Future therapeutic strategies may also aim at generating and using specifically "designed" blocking or agonist bacterial peptides that could possibly challenge and perhaps re-educate the mucosal immune system (Packey et al. 2008).

#### 6.1.1.2 The roles of PRMs in CD pathogenesis

Although over seven independent CD GWA studies and one meta-analysis have been published since the discovery of NOD2 association with CD, none of these GWAS have identified another PRM candidate. Our *NLRP3* association results consist of the first study implicating another PRM in the CD pathogenesis since the discovery of NOD2 in 2001. Such discoveries are important for patient care because it allow us to gain more understanding of key pathways and pathogenic mechanisms involved in the CD pathogenesis, such as the importance of intra-cellular bacterial sensors in this case. Our *NLRP3* association results further supports the immunopathogenesis hypothesis of CD, stating that it results from an inappropriate mucosal immune response towards the microbial flora in a genetically susceptible host.

Interestingly, we observed no significant associations between the *NLRP3* variants and UC, a result similar to what has been previously reported for the *NOD2* mutations and the autophagy genes *IRGM* and *ATG16L1*, which have been associated specifically to CD (Cho 2008; Fisher et al. 2008; Franke et al. 2008b). Together, these results suggest that alterations in the intracellular sensing and processing of bacteria may constitute a central feature specific to the pathogenesis

of CD, and not to UC. The fact that CD and UC do not fully share these associations may indicate distinct underlying mechanisms in these diseases.

We now know that PRMs like NOD2 and NALP3 are potentially involved in the initiation phase of CD, and it is likely that future studies will also implicate other PRMs. Despite the advance of knowledge in the field of innate immunity and in the understanding of the way harmful pathogens are detected, several questions remain unanswered. Indeed, it is still unclear how the bacterial products are delivered to the cytosol, and we don't know whether these PRMs bind directly or indirectly to their cognate agonists. Moreover, little is known about the mechanisms employed by these proteins to eliminate intracellular pathogens. Hence, although the genetic and first functional studies point towards an important role played by these PRMs in the onset of CD, additional biological research is still needed to further elucidate the way these receptors function and contribute to CD pathogenesis.

#### 6.1.1.3 Follow-up I: further dissection of the NALP3 inflammasome

Our *NLRP3* results indirectly implicate the NALP3 inflammasome pathway in the CD pathophysiology. However, subsequent evaluation of the *MEFV* gene, encoding the pyrin protein that also interact with the NALP3 inflammasome pathway, has excluded the role of this gene in CD and UC pathogenesis. *NLRP3-MEFV* gene-gene interaction analyses reported in Chapter 4 did not bring further insight into a potential involvement of *MEFV* in IBD susceptibility. Although this was beyond the scope of these projects, it would be interesting to screen all the components involved in the NALP3 inflammasome

pathway, such as ASC, CARDINAL, caspase-1, IL-1$\beta$, and see whether gene-gene interactions would further implicate this signaling cascade in the CD pathogenesis.

#### 6.1.1.4 Follow-up II: Evaluation of other genes responsible for HPFS

In addition to being involved in intracellular bacterial sensing and in the inflammasome pathway, mutations in the two genes on which Chapters 2-4 focus (*NLRP3* and *MEFV*) are also associated with the development of four HPFS, namely MWS, FCAS, NOMID (all three referred to as CAPS), and FMF (Ryan et al. 2008a). Indeed, these two genes are also known to play a central role in the regulation of fever, which is one of the most primitive defense mechanisms that exist in humans to fight the surrounding pathogenic bacteria. Our *NRLP3* results confirm one of our initial hypotheses, stating that while severe mutations in these genes will cause the respective severe monogenic fever syndrome, milder genetic changes such as SNPs, may give rise to more subtly altered protein function, and hence contribute to the basis of chronic inflammatory diseases, such as CD and UC. To further explore this latter hypothesis, it would be interesting to evaluate whether genes responsible for other HPFS, such as the tumor necrosis factor (TNF) receptor-associated periodic syndrome (TRAPS) and the hyperimmunoglobulinemia D with periodic fever syndrome (HIDS), also contribute to IBD susceptibility.

## 6.1.2 PI-IBS & IBS immunopathogenesis

### 6.1.2.1 PI-IBS association results in perspectives

The unfortunate Walkerton tragedy provided a unique opportunity to study the genetic determinants of post-infectious irritable bowel syndrome (PI-IBS) in a large, well-defined study cohort with simultaneous and well-characterized exposure. The identification of genetic factors involved in PI-IBS is important as it provides new insight into the pathophysiology of this disorder. Previous IBS genetic studies have proven rather challenging because of the heterogeneity of the disease phenotype and sample ethnicity, small sample size of cohorts, and lack of rigorous clinical and epidemiological data. Hitherto, no gene had yet been confirmed to be involved in the development of IBS-like symptoms. In this study, we had the advantage of working with a more homogeneous population cohort derived from a single community, where the environmental trigger was known and where we had access to a wide-range of clinical and epidemiological data.

Chapter 5 consists of the first genetic study ever conducted on PI-IBS patients. We identified three genes, namely *TLR9*, *CDH1*, and *IL6*, which all contribute to PI-IBS susceptibility. These observations are consistent with current conceptual models of IBS pathogenesis, and emphasize the important roles of gut flora, intestinal epithelial barrier function, and inflammatory pathways. Based on our association results, **Figure 6.1** gives a simplistic overview of the working hypothesis.

The observations that genetic risk factors contribute to the development of IBS following an episode of acute gastroenteritis may lead in the future to improved risk stratification and treatment following outbreaks.

### 6.1.2.2 Working hypothesis

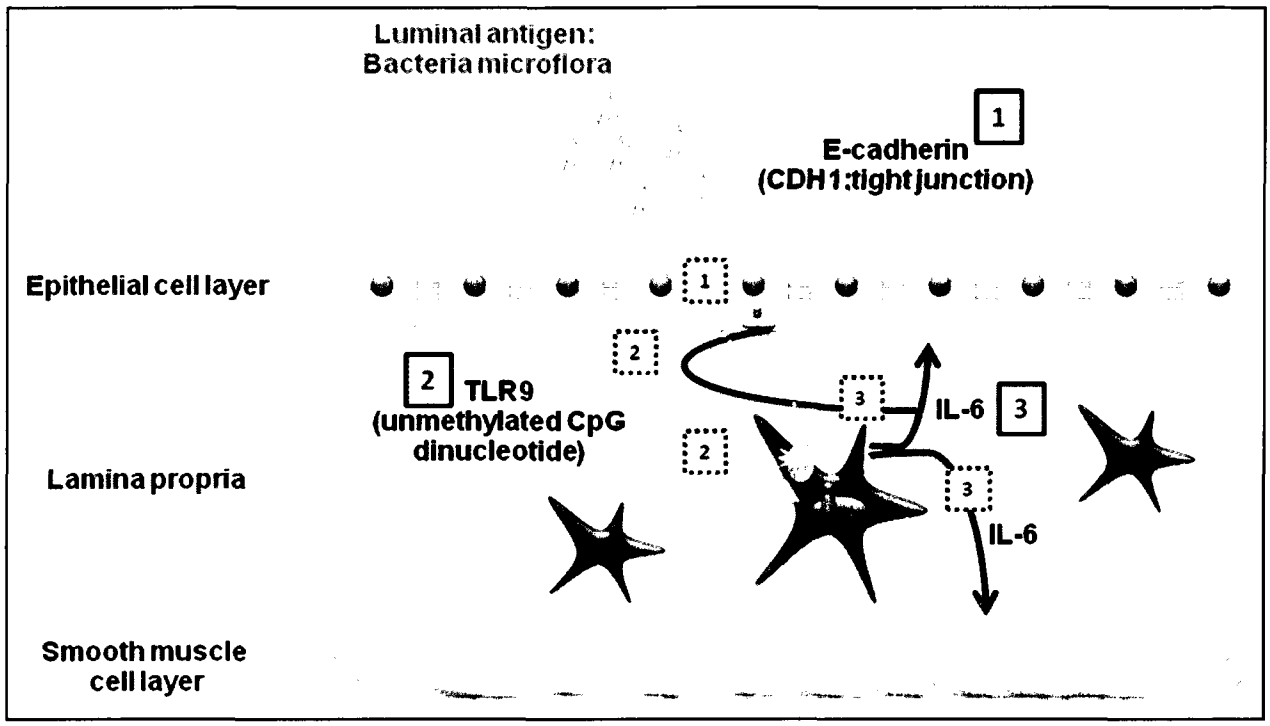



**Figure 6.1: Walkerton PI-IBS association results: working hypothesis**

Based on the known biological functions of the candidate genes associated with PI-IBS susceptibility, we can make the following working hypothesis to integrate the results: 1) the presence of enteric pathogens in the GI tract, such as *Escherichia Coli* and *Campylobacter jejuni* both present in the Walkerton outbreak, are known to have devised several ways to increase mucosal permeability, mainly through disruption epithelial tight junctions (TJ) like CDH1. If CDH1 is not properly up-regulated because of genetic predispositions, it may contribute in further weakening this TJ, leading to the increase in intestinal permeability that is observed in PI-IBS patients. 2) Subsequently, if TRL9 expression is not properly up-regulated when the intestinal epithelial and immune effector cells are exposed and challenged with enteric pathogens, resulting from a defect in properly recognizing the initial threat, it could contribute to the defective innate immune response documented in those who develop PI-IBS, 3) which could be further enhanced through inappropriate regulation of IL-6 production, leading to an impairment of pathogen clearance, and therefore to higher local antigen exposure that may contribute to further activating the intestinal immune system of PI-IBS patients, providing a mechanism for perpetuating inflammation.

#### 6.1.2.3 Challenges of replicating and validating the PI-IBS association results

The use of this unique Walkerton PI-IBS population cohort is a double-edged sword as no other well-characterized cohort of this size exists worldwide, and thus, we do not have the possibility of replicating these results. However, one of the greatest significance of the results obtained from a genetic study that used a more homogeneous PI-IBS patient sample set is that it will enable the formulation of novel genetic risk factors hypotheses that could also contribute to the development of IBS without an infectious onset. Hence, a next step would be to assess whether our findings could apply equally as well to this class of IBS patients.

For many years, IBS was not considered as a disease entity as no specific physiological and psychological abnormality had been found to be absolutely indicative of this disorder (Ohman et al. 2007). The precise molecular mechanism(s) underlying IBS remain essentially unknown, and the research in the fields of IBS and PI-IBS are still at a stage of infancy. The prospect of identifying a genetic cause for IBS is important, because it will raise the possibility of further confirming that IBS is a disease entity, suggest new insight into the pathophysiology of the disorder, and provide new targets for drug development.

It is unlikely that all patients with IBS-like symptoms have the same pathophysiology, and as research progresses, we should be able to further characterize and classify these patients; a better characterization and classification of patients will contribute towards building more homogenous cohorts, which

should help in uncovering more easily the genetic risk factors involved in the disease. A simple example comes from the sub-classification of IBS patients based on their predominant stool consistency; it is very unlikely that the biological pathways underlying the symptoms in a diarrhea-prone and a constipation-prone IBS patient will be the same. Given that a cohort composed of IBS patients with no infectious onset will be likely more heterogeneous in terms of clinical presentations than one composed of PI-IBS patients, it will be important to well-select the patients for subsequent validation of our results. For a start, since the most prominent bowel habit in PI-IBS patients is diarrhea, validation effort should focus on IBS patients presenting with this stool consistency. Overall, a better characterization of patient subgroups with regard to clinical, personal, and environmental factors is needed for genetic studies to be more successful in the future.

#### 6.1.2.4 Could IBS be an inflammatory disorder?

As explained in Chapter 5, our association results are in keeping with the pathophysiologic changes observed in patients with PI-IBS, which include increased intestinal permeability, low grade inflammation, and immune activation in the colonic mucosa. It has been hypothesized that a low-grade mucosal inflammatory process, albeit undetectable endoscopically or with conventional mucosal histology, may also play a role in IBS pathogenesis, implying that in at least a subpopulation of patients, IBS could be an inflammatory disorder. We would thus be more likely to validate our association results in this subpopulation

of patients. Four lines of evidence support this latter hypothesis. First, as explained previously, IBS symptoms may develop following an acute episode of infectious gastroenteritis (Spiller 2003a). Second, an increased number of inflammatory cells (e.g. mast cells, T lymphocytes, macrophages) have been detected in the colonic and ileal mucosa as well as in the muscularis externa of jejunum of patients with IBS. (Barbara et al. 2004a). Third, inflammatory cells in the intestinal mucosa of IBS patients have been reported to be activated (Chadwick et al. 2002; Barbara et al. 2004b) and to release a number of inflammatory mediators (Barbara et al. 2004a); these latter mediators are known to signal to epithelial, neuronal, and muscle cells leading to intestinal dysfunction (Barbara et al. 2006b). And last, patients in remission from IBD develop IBS-like symptoms with a higher than expected prevalence (Talley 2006; Burgmann et al. 2006). The prevalence of IBS symptoms in IBD patients in long-standing remission is 2 to 3 times higher than in normal population (Simrén et al. 2002).

We also need to keep in mind the possibility that the inflammatory components involved could be different in PI-IBS relative to IBS without an onset related to infection. If this is the case, it will be unlikely that *TLR9*, *IL6*, and *CDH1* genes play a role in IBS susceptibility. It is already known, for example, that patients with PI-IBS show a mucosal immune response that comprises predominantly CD4+T lymphocytes (Spiller et al. 2000) and cytokines such as IL1-$\beta$ (Gwee et al. 2003), whereas IBS patients without such a history have a response characterized primarily by infiltration of mast cells (Barbara et al. 2006b). Additional characterization of the inflammatory components implicated in the development of PI-IBS and IBS is needed to draw parallels between these

two disorders and to determine whether or not the components involved are distinct.

#### 6.1.2.5 PI-IBS and IBS genetic architecture & future work

The genetic variants conferring risk to PI-IBS and IBS may be commonly present in the general population, and each genetic variant may potentially only weakly contribute to the pathogenesis and symptoms, making them hard to uncover. Efforts in the near future should first focus on building cohorts of patients as homogeneous as possible to increase the chance of identifying the genetic risk factors involved. Although our association results consist of a first step towards a better understanding of the factors involved in the development of PI-IBS, we recognize that the study of genetic risk factors implicated in PI-IBS and IBS susceptibility is still at the stage of infancy and that many more genes are likely to be discovered in the coming years. Because of budgetary constrains, we have opted for candidate gene approach, looking specifically at functional variants previously associated in these genes. Hence, our study can only exclude the involvement of the reported functional and tagging SNPs that we have evaluated in PI-IBS, as well as SNPs in high LD and tagged by them. However, as mentioned in Chapter 5, we cannot exclude the possible involvement of the candidate genes evaluated in the exploratory phase unless thorough fine mapping experiments are performed in all of these regions. With access to additional funding, the richness and uniqueness of the PI-IBS Walkerton population cohort

should be further exploited to uncover additional genetic risk factors involved by the use of larger scale approaches, such as the GWA study strategy.

## 6.2 The genetic architecture of IBD

The meta-analysis by Barrett and colleagues (Barrett et al. 2008) estimated that well-established associations with CD account for only a fifth of the genetic variance observed in CD. Hence, most of the inherited components of the disease predisposition remain essentially unexplained. These estimations have not yet been published for UC, but they are likely to be low as well. One question may arise: where is the remaining genetic variance contributing to CD and UC susceptibility?

### 6.2.1 Limitations and challenges of IBD GWAS

Many more IBD loci remain to be identified by GWAS. When interpreting this ratio, several factors need to be taken into consideration.

#### 6.2.1.1 Tagging SNPs vs. causal SNPs

First, most of the SNPs that were genotyped in these GWAS were selected to tag the genome efficiently, rather than for their possible effect on gene function. Consequently, most CD-associated SNPs identified so far are unlikely to be the causal variants that actually confer disease susceptibility. The *IRGM* locus association with CD is a good example; it was originally identified on the basis of genome-wide SNP association analysis and a subsequent study by

McCaroll *et al.* identified the causal variant to be a deletion upstream of the promoter, affecting tissuse-specific expression (McCaroll et al. 2008). Hence, the regions associated with CD and UC that were uncovered through the numerous GWA have yet to be further scrutinized through fine-mapping and resequencing approaches to identify the specific genes and variants responsible for the association signal. Overall, each locus-attributable risk factor will likely be higher than the currently estimated one as we identify the causal risk factor(s) in each associated regions.

#### 6.2.1.2 Independent common risk variant(s)

Moreover, it is also important to remember that a single associated locus may contain multiple independent common risk variants. This was the case for the *IL23R* gene association with CD, where logistic regression showed that the identified coding SNP (Arg381Gln) was insufficient to explain the association signal, and that other *IL23R* variants as well as variants the intergenic region between *IL23R* and *IL12RB2* showed additional independent associations (Duerr et al. 2006). These unexplored potential independent causal alleles may also account for additional genetic risk.

#### 6.2.1.3 Population studied

Furthermore, the great majority of published studies have used populations of European ancestry. Larger, more comprehensive, and more diverse GWAS should be able to reveal many more loci.

#### 6.2.1.4 Statistical power

Additionally, studies reported to date had very low statistical power and thus must have missed many loci with common variants of similar or smaller effects. Indeed, given the modest effect sizes of common variants now known to exist, which in average for CD consists of a factor <2 per alleles (Barrett et al. 2008), and the need to exceed stringent statistical thresholds, the first waves of GWAS provided low power to discover disease-causing loci (Altshuler et al. 2007). It is also important to keep in mind that studies of hundreds to several thousands cases have been necessary to identify only a fifth of the loci contributing to CD susceptibility (Barrett et al. 2008). For many of the identified loci, chance in the form of sampling error, the so called "winner's curse", probably played a role in boosting their detectability (Göring et al. 2001; Lohmueller et al. 2003). Therefore, many loci of equivalent effect size are likely to also confer susceptibility and are simply unidentified for the moment.

#### 6.2.1.5 Assessing structural and rare variants

Moreover, it is possible that some disease loci will contain only structural or rare variants. However, the first published GWAS did not use technologies that comprised proxies for common structural variants, such as CNV (copy number variation). Moreover, the available commercial array used for GWAS are designed to provide excellent coverage of common SNPs, but have very limited potential to capture rare and low frequency variants with MAF lower than 5% (Zeggini et al. 2005). Systematic resequencing of the entire associated region, at

least to the point at which LD substantially decays, in large sample sets will be required to identify these rare variants, like it is currently being done for the 1000 genome project (Kuehn 2008), which is creating a new map of the human genome by sequencing the genomes of at least 1000 individuals from around the world (www.1000genomes.org). In addition to uncover rare associated variants, another important benefit of the deep-sequencing approach is the potential to uncover alleles with more severe effects at the molecular and clinical levels; such variants are likely to be more attractive substrates for functional and clinical investigation that the common variants originally identified. As the available sequencing technologies will improve, this task will become more easily achievable than previously (Mardis 2008). The extent to which low frequency and structural variants will contribute to disease predisposition represents one of the major unanswered questions in IBD genetics. Such variants could, individually and collectively, have greater impact in terms of explaining disease risk, than the common variants that have been identified so far using the GWAS approach (McCarthy et al. 2008).

## 6.2.2 Other challenges in IBD genetics

### 6.2.2.1 Gene-gene and gene-environment interactions

Almost all GWAS published to date have focused on detecting and characterizing the main genetic effects. However, gene-gene and gene-environment interactions may also play important role in disease risk. Although searches have not yet found any evidences of epistasis in CD and UC, it may

simply reflect the limited power to assess the many possible modes of interactions and thresholds of effects. Indeed, **1)** one of the important limitations is the computational burden imposed by any comprehensive search for higher-order effects of gene-gene interactions. Exhaustive studies of gene-gene effects are highly dependent on knowledge of the exact causal variants, larger data-sets, and more efficient computational approaches. **2)** The challenge with respect of identifying gene-environment is even greater, as many of the parameters that are likely to be relevant to disease predisposition are difficult to measure in a detailed and standardized manner in large sample collections.

Hence, several obstacles need be overcome to assemble the appropriate tools in order to correctly assess the potential contribution of gene-gene and gene-environment interactions to IBD susceptibility. Once these patterns of associations and interactions are understood, effects of specific gene and environmental exposures on each phenotype will likely be larger.

#### 6.2.2.2 Epigenetic contribution(s)

Additional potential sources of heritability contributing to disease predisposition may also come from epigenetic effects, such as methylation or histone modifications. Indeed, epigenetic changes that are transmitted across generations, but where the inherited epigenetic state does not track with underlying DNA sequence variations, might have been missed in previous GWAS. Consequently, high-throughput tools need to be developed to assess the contribution of such epigenetic effects in IBD susceptibility.

### 6.2.3 IBD GWAS results: the beginning of a journey

Overall, although the proportion of genetic variance explained is certain to grow in the coming years, we also need to keep in mind that to the extent that CD and UC are influenced by tiny effects of hundreds of loci or highly heterogeneous rare mutations, it may be impractical to assemble sufficiently large samples to give a complete account of the genetic risk factors involved. We can conclude that, although substantial efforts were gathered to conceive these numerous GWA studies (see **Tables 1.1-1.2**), the first wave of results represent only the starting point on the journey to elucidating and understanding the genetic basis of IBD, and translating this knowledge into clinically useful insights.

## 6.3 Personalized medicine: are we there yet?

The advances in the field of genomic research have greatly improved the progress and speed at which we now discover disease susceptibility genes, and have fueled expectations about the potential opportunities of using genetic profiling as a tool for personalized medicine. Most of the current therapies used to treat CD and UC rely on non-specific immune-suppressive approaches, in part due to a lack of knowledge of the specific mechanisms that cause disease onset and progress. Understanding the genetic factors contributing to these diseases is therefore likely to have a significant effect on public health. Genetic factors are also thought to have a predominant role in determining progression to clinical disease, and identifying these factors might contribute to the development of targeted preventive therapies. Indeed, there are now hopes in the IBD scientific

community that the genetic field will allow the identification of individuals at risk of developing CD or UC by defining specific genetic backgrounds at risk (i.e. disease prevention).

The use of personalized medicine, such as genetic profiling, requires a test that fairly accurately predicts disease risk, particularly when it concerns decisions about therapeutic interventions, or interventions that are invasive, expensive, or have major side effects. Although novel gene discoveries reported by GWA studies will certainly improve the prediction of IBD onset, the question is whether this improvement will be sufficient to enable personalized medicine.

### 6.3.1 Challenges in personalized medicine

Personalized medicine is very straightforward for monogenic disorders, which are completely or predominantly caused by DNA variations in one single gene, as in the case of HPFS. The development of genetic testing on such genetic basis for informing individuals at risk about their future health status and for deciding upon specific interventions can thus be done more easily. However, the etiology of complex traits, like CD and UC, which result from a complex interplay of genetic and environmental factors where each factor may only have a minor contribution to the occurrence of the disease, makes it immensely more complicated to develop accurate prediction models to be used for personalized medicine.

Although it is tempting to think that some genotypes will be able to predict the development of disease, the clinical outcomes or the response to drugs,

accurate and precise predictions will only be possible when we will be able to understand the essential genetic and environmental factors involved in the pathogenic mechanisms underlying CD and UC development.

### 6.3.2 The future of personalized medicine in IBD

At this point of knowledge, with only 20% of CD heritability to be accounted for (Barrett et al. 2008), we still have very limited understanding of the exact causal variants, and thus the available disease models are likely to be of poor predictability. As explained previously, many more common variants also contributing to CD and UC susceptibility remain to be uncovered; we cannot exclude for the moment the involvement of rare variants with major effects, we have limited insights in ethnic differences, gene expression and functions, as well as in gene-gene and gene-environment interactions that could all contribute to disease development. Since our understanding of CD and UC pathogenesis will only accelerate in the coming year, resolving many of these issues, this should lead to improved genetic prediction modeling. The extent to which the uncovered genetic associations will figure in personalized medicine will depend on the level of accuracy that the prediction models will be able to attain, and on whether there will be intervention whose effectiveness will be improved by the knowledge of a genetic test. In the coming years, the great challenge will thus be to integrate these results and build more accurate disease prediction models with the hopes of eventually translating the emerging genomic knowledge into public health and medical care.

# 7 CONCLUSION

Over the course of this PhD project, we used a hypothesis-driven candidate approach to try to understand the contribution of innate immunity risk candidates to the susceptibility of CD, UC, and PI-IBS. Our results have deepened our comprehension of the complex interrelationship between enteric infection, intestinal microbial flora, genetic susceptibility, and disease pathogenesis by providing novel insights about potential mechanisms involved in the development of these diseases. Nonetheless, future genetic, epidemiologic, animal model, and in-depth functional studies are needed to further evaluate the implication of the pathways that were highlighted by our studies, being the NALP3 inflammasome in the case of CD, and the contribution of *CDH1*, *TLR9*, and *IL6* in the case of PI-IBS. Knowledge of these disease pathways should not be limited to the causal genes and associated variants reported in this Ph.D. thesis; these signaling cascades should be further dissected as they may eventually lead to alternative strategies for prevention, diagnosis, and therapy for CD, UC, PI-IBS, and perhaps, to other inflammatory diseases. Finally, although uncovering new susceptibility genes is important as it may lead to new biological insights on the mechanisms underlying these diseases' pathogenesis, we need to keep in mind that these traits are all powerfully shaped by the environment, and that the solutions to several of the important problems encountered with these diseases are likely to lie outside our genes.

# 8 REFERENCES

Aamodt G, Jahnsen J, Bengtson MB, Moum B, Vatn MH; IBSEN Study Group. Geographic distribution and ecological studies of inflammatory bowel disease in southeastern Norway in 1990-1993. Inflamm Bowel Dis 2008; 14: 984-991.

Abraham C, Cho JH. IL-23 and Autoimmunity: New Insights into the Pathogenesis of Inflammatory Bowel Disease. Annu Rev Med 2009; 60: 97-110.

Adam B, Liebregts T, Gschossmann JM, Krippner C, Scholl F, Ruwe M, Holtmann G. Severity of mucosal inflammation as a predictor for alterations of visceral sensory function in a rat model. Pain 2006; 123(1-2): 179-186.

Aganna E, Martinon F, Hawkins PN, Ross JB, Swan DC, Booth DR, Lachmann HJ, Bybee A, Gaudet R, Woo P, Feighery C, Cotter FE, Thome M, Hitman GA, Tschopp J, McDermott MF. Association of mutations in the NALP3/CIAS1/PYPAF1 gene with a broad phenotype including recurrent fever, cold sensitivity, sensorineural deafness, and AA amyloidosis. Arthritis Rheum 2002; 46: 2445-2452.

Agostini L, Martinon F, Burns K, McDermott MF, Hawkins PN, Tschopp J. NALP3 forms an IL-1beta-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. Immunity 2004; 20: 319-325.

Akiho H, Deng Y, Blennerhassett P, Kanbayashi H, Collins SM. Mechanisms underlying the maintenance of muscle hypercontractility in a model of postinfective gut dysfunction. Gastroenterology 2005; 129: 131-141.

Akira S. Toll-like receptors and innate immunity. Adv Immunol 2001; 78: 1-56.

Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell 2006; 124: 783-801.

Aksentijevich I, Torosyan Y, Samuels J, Centola M, Pras E, Chae JJ, Oddoux C, Wood G, Azzaro MP, Palumbo G, Giustolisi R, Pras M, Ostrer H, Kastner DL. Mutation and haplotype studies of familial Mediterranean fever reveal new ancestral relationships and evidence for a high carrier frequency with reduced penetrance in the Ashkenazi Jewish population. Am J Hum Genet 1999; 64: 949-962.

Aksentijevich I, Nowak M, Mallah M, Chae JJ, Watford WT, Hofmann SR, Stein L, Russo R, Goldsmith D, Dent P, Rosenberg HF, Austin F, Remmers EF, Balow JE Jr, Rosenzweig S, Komarow H, Shoham NG, Wood G, Jones J, Mangra N, Carrero H, Adams BS, Moore TL, Schikler K, Hoffman H, Lovell DJ, Lipnick R, Barron K, O'Shea JJ, Kastner DL, Goldbach-Mansky R. De novo CIAS1 mutations, cytokine activation, and evidence for genetic heterogeneity in patients with neonatal-onset multisystem inflammatory disease (NOMID): a new member of the expanding family of pyrin-associated autoinflammatory diseases. Arthritis Rheum 2002; 46: 3340-3348.

Aldea A, Calafell F, Arostegui JI, Lao O, Rius J, Plaza S, Masó M, Vives J, Buades J, Yagüe J. The west side story: MEFV haplotype in Spanish FMF

patients and controls, and evidence of high LD and a recombination "hot-spot" at the MEFV locus. Hum Mutat 2004; 23: 399.

Andresen V, Camilleri M, Kim HJ, Stephens DA, Carlson PJ, Talley NJ, Saito YA, Urrutia R, Zinsmeister AR. Is there an association between GNbeta3-C825T genotype and lower functional gastrointestinal disorders? Gastroenterology 2006; 130: 1985-1994.

Anderson JP, Mueller JL, Rosengren S, Boyle DL, Schaner P, Cannon SB, Goodyear CS, Hoffman HM. Structural, expression, and evolutionary analysis of mouse CIAS1. Gene 2004; 338: 25-34.

Anderson KV, Bokla L, Nüsslein-Volhard C. Establishment of dorsal-ventral polarity in the Drosophila embryo: the induction of polarity by the Toll gene product. Cell 1985; 42: 791-798.

Annese V, Valvano MR, Palmieri O, Latiano A, Bossa F, Andriulli A. Multidrug resistance 1 gene in inflammatory bowel disease: a meta-analysis. World J Gastroenterol 2006; 12: 3636-3644.

Artis D. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. Nat Rev Immunol 2008; 8: 411-420.

Altshuler D, Daly M. Guilt beyond a reasonable doubt. Nat Genet 2007; 39: 813-815.

Aujla SJ, Dubin PJ, Kolls JK. Th17 cells and mucosal host defense. Semin Immunol 2007; 19: 377-382.

Auron PE, Webb AC, Rosenwasser LJ, Mucci SF, Rich A, Wolff SM, Dinarello CA. Nucleotide sequence of human monocyte interleukin 1 precursor cDNA. Proc Natl Acad Sci U S A. 1984; 81: 7907-7911.

Ausubel FM. Are innate immune signaling pathways in plants and animals conserved? Nat Immunol 2005; 6: 973-979.

Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci USA 2004; 101: 15718-15723.

Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005; 307: 1915-1920.

Baldwin AS Jr. Series introduction: the transcription factor NF-kappaB and human disease. J Clin Invest 2001; 107: 3-6.

Barbara G, De Giorgio R, Stanghellini V, Cremon C, Salvioli B, Corinaldesi R. New pathophysiological mechanisms in irritable bowel syndrome. Aliment Pharmacol Ther 2004a; 20 (Suppl 2):1-9.

Barbara G, Stanghellini V, De Giorgio R, Cremon C, Cottrell GS, Santini D, Pasquinelli G, Morselli-Labate AM, Grady EF, Bunnett NW, Collins SM, Corinaldesi R. Activated mast cells in proximity to colonic nerves correlate with abdominal pain in irritable bowel syndrome. Gastroenterology 2004b; 126: 693-702.

Barbara G, Stanghellini V, Brandi G, Cremon C, Di Nardo G, De Giorgio R, Corinaldesi R. Interactions between commensal bacteria and gut sensorimotor function in health and disease. Am J Gastroenterol 2005; 100: 2560-2568.

Barbara G. Mucosal barrier defects in irritable bowel syndrome. Who left the door open? Am J Gastroenterol 2006a; 101:1295-1298.

Barbara G, Stanghellini V, De Giorgio R, Corinaldesi R. Functional gastrointestinal disorders and mast cells: implications for therapy. Neurogastroenterol Motil 2006b; 18(1): 6-17.

Barbara G, Stanghellini V, Cremon C, De Giorgio R, Corinaldesi R. Almost all irritable bowel syndromes are post-infectious and respond to probiotics: controversial issues. Dig Dis 2007a; 25: 245-248.

Barbara G, Wang B, Stanghellini V, de Giorgio R, Cremon C, Di Nardo G, Trevisani M, Campi B, Geppetti P, Tonini M, Bunnett NW, Grundy D, Corinaldesi R. Mast cell-dependent excitation of visceral-nociceptive sensory neurons in irritable bowel syndrome. Gastroenterology 2007b; 132: 26-37.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21: 263-265.

Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhart AH, Targan SR, Xavier RJ; NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi

J, Mathew CG, Parkes M, Georges M, Daly MJ. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008; 40: 955-962.

Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. Lancet 2007; 369: 1627-1640.

Beaudet AL, Belmont JW. Array-based DNA diagnostics: let the revolution begin. Annu Rev Med 2008; 59:113-129.

Belkhadir Y, Subramaniam R, Dangl JL. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Curr Opin Plant Biol 2004; 7: 391-399.

Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, Kopla R, Modica F, Pohl M, Varde S, Zhao R, Zhao X, Boyce-Jacino MT, Yassen A. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques 2002; Suppl: 70-72, 74, 76-77.

Bengtson MB, Rønning T, Vatn MH, Harris JR. Irritable bowel syndrome in twins: genes and environment. Gut 2006; 55: 1754-1759.

Benno Y, Sawada K, Mitsuoka T. The intestinal microflora of infants: composition of fecal flora in breast-fed and bottle-fed infants. Microbiol Immunol 1984; 28: 975-986.

Bercík P, Wang L, Verdú EF, Mao YK, Blennerhassett P, Khan WI, Kean I, Tougas G, Collins SM. Visceral hyperalgesia and intestinal dysmotility in a mouse model of postinfective gut dysfunction. Gastroenterology 2004; 127: 179-187.

Bercik P, Verdu EF, Collins SM. Is irritable bowel syndrome a low-grade inflammatory bowel disease? Gastroenterol Clin North Am 2005; 34: 235-45, vi-vii.

Bernot A, da Silva C, Petit JL, Cruaud C, Caloustian C, Castet V, Ahmed-Arab M, Dross C, Dupont M, Cattan D, Smaoui N, Dodé C, Pêcheux C, Nédelec B, Medaxian J, Rozenbaum M, Rosner I, Delpech M, Grateau G, Demaille J, Weissenbach J, Touitou I. Non-founder mutations in the MEFV gene establish this gene as the cause of familial Mediterranean fever (FMF). Hum Mol Genet 1998; 7: 1317-1325.

Bertin J, DiStefano PS. The PYRIN domain: a novel motif found in apoptosis and inflammation proteins. Cell Death Differ 2000; 7: 1273-1274.

Bettelli E, Korn T, Oukka M, Kuchroo VK. Induction and effector functions of T(H)17 cells. Nature 2008; 453: 1051-1057.

Bhat A, Naguwa SM, Gershwin ME. Genetics and new treatment modalities for familial Mediterranean fever. Ann N Y Acad Sci 2007; 1110: 201-208.

Birmingham CL, Smith AC, Bakowski MA, Yoshimori T, Brumell JH. Autophagy controls Salmonella infection in response to damage to the Salmonella-containing vacuole. J Biol Chem 2006; 281: 11374-11383.

Bisping G, Lügering N, Lütke-Brintrup S, Pauels HG, Schürmann G, Domschke W, Kucharzik T. Patients with inflammatory bowel disease (IBD) reveal increased induction capacity of intracellular interferon-gamma (IFN-gamma) in peripheral CD8+ lymphocytes co-cultured with intestinal epithelial cells. Clin Exp Immunol 2001; 123:15-22.

Black RA, Kronheim SR, Sleath PR. Activation of interleukin-1 beta by a co-induced protease. FEBS Lett 1989; 247: 386-390.

Bommelaer G, Poynard T, Le Pen C, Gaudin AF, Maurel F, Priol G, Amouretti M, Frexinos J, Ruszniewski P, El Hasnaoui A. Prevalence of irritable bowel syndrome (IBS) and variability of diagnostic criteria. Gastroenterol Clin Biol 2004; 28 (6-7 Pt 1): 554-561.

Boneca IG. The role of peptidoglycan in pathogenesis. Curr Opin Microbiol 2005; 8: 46-53.

Borgaonkar MR, Ford DC, Marshall JK, Churchill E, Collins SM. The incidence of irritable bowel syndrome among community subjects with previous acute enteric infection. Dig Dis Sci. 2006; 51: 1026-1032.

Bouma G, Xia B, Crusius JB, Bioque G, Koutroubakis I, Von Blomberg BM, Meuwissen SG, Peña AS. Distribution of four polymorphisms in the tumour necrosis factor (TNF) genes in patients with inflammatory bowel disease (IBD). Clin Exp Immunol 1996; 103: 391-396.

Bouma G, Strober W. The immunological and genetic basis of inflammatory bowel disease. Nat Rev Immunol 2003; 3: 521-533.

Bousvaros A, Guandalini S, Baldassano RN, Botelho C, Evans J, Ferry GD, Goldin B, Hartigan L, Kugathasan S, Levy J, Murray KF, Oliva-Hemker M, Rosh JR, Tolia V, Zholudev A, Vanderhoof JA, Hibberd PL. A randomized, double-blind trial of Lactobacillus GG versus placebo in addition to standard maintenance therapy for children with Crohn's disease. Inflamm Bowel Dis 2005; 11: 833– 839.

Bradesi S, McRoberts JA, Anton PA, Mayer EA. Inflammatory bowel disease and irritable bowel syndrome: separate or unified? Curr Opin Gastroenterol 2003;19: 336-342.

Browning BL, Huebner C, Petermann I, Demmers P, McCulloch A, Gearry RB, Barclay ML, Shelling AN, Ferguson LR. Association of DLG5 variants with inflammatory bowel disease in the New Zealand Caucasian population and meta-analysis of the DLG5 R30Q variant. Inflamm Bowel Dis 2007; 13: 1069-1076.

Burgmann T, Clara I, Graff L, Walker J, Lix L, Rawsthorne P, McPhail C, Rogala L, Miller N, Bernstein CN. The Manitoba Inflammatory Bowel Disease Cohort Study: prolonged symptoms before diagnosis--how much is irritable bowel syndrome? Clin Gastroenterol Hepatol 2006; 4: 614-620.

Burns KA, Martinon F. Inflammatory diseases: is ubiquitinated NEMO at the hub? Curr Biol 2004; 14: R1040-1042.

Cain KC, Headstrom P, Jarrett ME, Motzer SA, Park H, Burr RL, Surawicz CM, Heitkemper MM. Abdominal pain impacts quality of life in women with irritable bowel syndrome. Am J Gastroenterol 2006; 101: 124-132.

Camilleri M, Northcutt AR, Kong S, Dukes GE, McSorley D, Mangel AW. Efficacy and safety of alosetron in women with irritable bowel syndrome: a randomised, placebo-controlled trial. Lancet 2000; 355: 1035-1040.

Camilleri M. Pharmacogenomics and serotonergic agents: research observations and potential clinical practice implications. Neurogastroenterol Motil. 2007; 19 (Suppl 2): 40-45.

Cantor MJ, Nickerson P, Bernstein C. The role of cytokine gene polymorphisms in determining disease susceptibility and phenotype in inflammatory bowel disease. Am J Gastroenterol 2005; 100: 1134-1142.

Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001; 2: 91-99.

Cargill M, Schrodi SJ, Chang M, Garcia VE, Brandon R, Callis KP, Matsunami N, Ardlie KG, Civello D, Catanese JJ, Leong DU, Panko JM, McAllister LB, Hansen CB, Papenfuss J, Prescott SM, White TJ, Leppert MF, Krueger GG, Begovich AB. A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. Am J Hum Genet 2007; 80: 273-290.

Cargnin F, Flora A, Di Lascio S, Battaglioli E, Longhi R, Clementi F, Fornasari D. PHOX2B regulates its own expression by a transcriptional auto-regulatory mechanism. J Biol Chem 2005; 280: 37439-37448.

Cario E, Becker A, Sturm A, Goebell H, Dignass AU. Peripheral blood mononuclear cells promote intestinal epithelial restitution in vitro through an interleukin-2/interferon-gamma-dependent pathway. Scand J Gastroenterol 1999; 34: 1132-1138.

Cario E. Innate immune signalling at intestinal mucosal surfaces: a fine line between host protection and destruction. Curr Opin Gastroenterol 2008; 24: 725-732.

Carneiro LA, Travassos LH, Girardin SE. Nod-like receptors in innate immunity and inflammatory diseases. Ann Med 2007; 39: 581-593.

Carneiro LA, Magalhaes JG, Tattoli I, Philpott DJ, Travassos LH. Nod-like proteins in inflammation and disease. J Pathol 2008; 214: 136-148.

Cash B, Sullivan S, Barghout V. Total costs of IBS: employer and managed care perspective. Am J Manag Care. 2005; 11(1 Suppl): S7-16.

Cattan D, Notarnicola C, Molinari N, Touitou I. Inflammatory bowel disease in non-Ashkenazi Jews with familial Mediterranean fever. Lancet 2000; 355: 378-379.

Centola M, Wood G, Frucht DM, Galon J, Aringer M, Farrell C, Kingma DW, Horwitz ME, Mansfield E, Holland SM, O'Shea JJ, Rosenberg HF, Malech HL, Kastner DL. The gene for familial Mediterranean fever, MEFV, is expressed in early leukocyte development and is regulated in response to inflammatory mediators. Blood 2000; 95: 3223-3231.

Chadwick VS, Chen W, Shu D, Paulus B, Bethwaite P, Tie A, Wilson I. Activation of the mucosal immune system in irritable bowel syndrome. Gastroenterology 2002; 122: 1778-1783.

Chae JJ, Centola M, Aksentijevich I, Dutra A, Tran M, Wood G, Nagaraju K, Kingma DW, Liu PP, Kastner DL. Isolation, genomic organization, and expression analysis of the mouse and rat homologs of MEFV, the gene for familial mediterranean fever. Mamm Genome 2000; 11: 428-435.

Chae JJ, Komarow HD, Cheng J, Wood G, Raben N, Liu PP, Kastner DL. Targeted disruption of pyrin, the FMF protein, causes heightened sensitivity to endotoxin and a defect in macrophage apoptosis. Mol Cell 2003; 11: 591-604.

Chae JJ, Wood G, Masters SL, Richard K, Park G, Smith BJ, Kastner DL. The B30.2 domain of pyrin, the familial Mediterranean fever protein, interacts directly with caspase-1 to modulate IL-1beta production. Proc Natl Acad Sci USA 2006; 103: 9982-9987.

Chae JJ, Wood G, Richard K, Jaffe H, Colburn NT, Masters SL, Gumucio DL, Shoham NG, Kastner DL. The familial Mediterranean fever protein, pyrin, is cleaved by caspase-1 and activates NF-kappaB through its N-terminal fragment. Blood. 2008; 112:1794-1803.

Chan I, Liu L, Hamada T, Sethuraman G, McGrath JA. The molecular basis of lipoid proteinosis: mutations in extracellular matrix protein 1. Exp Dermatol 2007; 16: 881-890.

Chang L, Toner BB, Fukudo S, Guthrie E, Locke GR, Norton NJ, Sperber AD. Gender, age, society, culture, and the patient's perspective in the functional gastrointestinal disorders. Gastroenterology. 2006; 130: 1435-1446.

Chang YC, Chen TC, Lee CT, Yang CY, Wang HW, Wang CC, Hsieh SL. Epigenetic control of MHC class II expression in tumor-associated macrophages by decoy receptor 3. Blood 2008; 111: 5054-5063.

Chaudhary NA, Truelove SC. The irritable colon syndrome. A study of the clinical features, predisposing causes, and prognosis in 130 cases. Q J Med 1962; 31: 307-322.

Chen JJ, Li Z, Pan H, Murphy DL, Tamir H, Koepsell H, Gershon MD. Maintenance of serotonin in the intestinal mucosa and ganglia of mice that

lack the high-affinity serotonin transporter: Abnormal intestinal motility and the expression of cation transporters. J Neurosci 2001; 21: 6348-6361.

Chen X, Bykhovskaya Y, Tidow N, Hamon M, Bercovitz Z, Spirina O, Fischel-Ghodsian N. The familial mediterranean fever protein interacts and colocalizes with a putative Golgi transporter. Proc Soc Exp Biol Med 2000; 224: 32-40.

Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. Nat Rev Immunol 2008; 8: 458-466.

Chu ZL, Pio F, Xie Z, Welsh K, Krajewska M, Krajewski S, Godzik A, Reed JC. A novel enhancer of the Apaf1 apoptosome involved in cytochrome c-dependent caspase activation and apoptosis. J Biol Chem 2001; 276: 9239-9245.

Church LD, Cook GP, McDermott MF. Primer: inflammasomes and interleukin 1beta in inflammatory disorders. Nat Clin Pract Rheumatol 2008; 4: 34-42.

Coates MD, Mahoney CR, Linden DR, Sampson JE, Chen J, Blaszyk H, Crowell MD, Sharkey KA, Gershon MD, Mawe GM, Moses PL. Molecular defects in mucosal serotonin content and decreased serotonin reuptake transporter in ulcerative colitis and irritable bowel syndrome. Gastroenterology 2004; 126:1 657-1664.

COGENT Study. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008; 40: 1426-1435.

Collazo CM, Yap GS, Sempowski GD, Lusby KC, Tessarollo L, Woude GF, Sher A, Taylor GA. Inactivation of LRG-47 and IRG-47 reveals a family of interferon gamma-inducible genes with essential, pathogen-specific roles in resistance to infection. J Exp Med 2001; 194: 181-188.

Colomar A, Marty V, Médina C, Combe C, Parnet P, Amédée T. Maturation and release of interleukin-1beta by lipopolysaccharide-primed mouse Schwann cells require the stimulation of P2X7 receptors. J Biol Chem 2003; 278: 30732-30740.

Communi D, Janssens R, Suarez-Huerta N, Robaye B, Boeynaems JM. Advances in signalling by extracellular nucleotides. the role and transduction mechanisms of P2Y receptors. Cell Signal 2000; 12: 351-360.

Conway KE, McConnell BB, Bowring CE, Donald CD, Warren ST, Vertino PM. TMS1, a novel proapoptotic caspase recruitment domain protein, is a target of methylation-induced gene silencing in human breast cancers. Cancer Res 2000; 60: 6236-6242.

Creagh EM, O'Neill LA. TLRs, NLRs and RLRs: a trinity of pathogen sensors that co-operate in innate immunity. Trends Immunol 2006; 27: 352-357.

Creed F, Ratcliffe J, Fernandez L, Tomenson B, Palmer S, Rigby C, Guthrie E, Read N, Thompson D. Health-related quality of life and health care costs in severe, refractory irritable bowel syndrome. Ann Intern Med 2001; 134(9 Pt 2): 860-868.

Cremonini F, Camilleri M, McKinzie S, Carlson P, Camilleri CE, Burton D, Thomforde G, Urrutia R, Zinsmeister AR. Effect of CCK-1 antagonist, dexloxiglumide, in female patients with irritable bowel syndrome: a

pharmacodynamic and pharmacogenomic study. Am J Gastroenterol. 2005; 100: 652-663.

Crohn BB, Ginzburg L, Oppenheimer GD. Landmark article Oct 15, 1932. Regional ileitis. A pathological and clinical entity. JAMA. 1984; 251: 73-79.

Croucher PJ, Mascheretti S, Hampe J, Huse K, Frenzel H, Stoll M, Lu T, Nikolaus S, Yang SK, Krawczak M, Kim WH, Schreiber S. Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. Eur J Hum Genet 2003; 11: 6-16.

Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser AL, Barnich N, Bringer MA, Swidsinski A, Beaugerie L, Colombel JF. High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. Gastroenterology 2004; 127: 412-421.

Day TG, Ramanan AV, Hinks A, Lamb R, Packham J, Wise C, Punaro M, Donn RP. Autoinflammatory genes and susceptibility to psoriatic juvenile idiopathic arthritis. Arthritis Rheum 2008; 58: 2142-2146.

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005; 37: 1217-1223.

de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD. A high-resolution HLA

and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 2006; 38: 1166-1172.

De Jager PL, Franchimont D, Waliszewska A, Bitton A, Cohen A, Langelier D, Belaiche J, Vermeire S, Farwell L, Goris A, Libioulle C, Jani N, Dassopoulos T, Bromfield GP, Dubois B, Cho JH, Brant SR, Duerr RH, Yang H, Rotter JI, Silverberg MS, Steinhart AH, Daly MJ, Podolsky DK, Louis E, Hafler DA, Rioux JD; Quebec IBD Genetics Consortium; NIDDK IBD Genetics Consortium. The role of the Toll receptor pathway in susceptibility to inflammatory bowel diseases. Genes Immun. 2007; 8: 387-397.

Delbridge LM, O'Riordan MX. Innate recognition of intracellular bacteria. Curr Opin Immunol 2007; 19: 10-16.

de Mesquita MB, Civitelli F, Levine A. Epidemiology, genes and inflammatory bowel diseases in childhood. Dig Liver Dis 2008; 40: 3-11.

Deretic V. Autophagy as an immune defense mechanism. Curr Opin Immunol 2006; 18: 375-382.

Derkx B, Taminiau J, Radema S, Stronkhorst A, Wortel C, Tytgat G, van Deventer S. Tumour-necrosis-factor antibody treatment in Crohn's disease. Lancet 1993; 342:173–174.

DeYoung BJ, Innes RW. Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol 2006; 7: 1243-1249.

Diaz A, Hu C, Kastner DL, Schaner P, Reginato AM, Richards N, Gumucio DL. Lipopolysaccharide-induced expression of multiple alternatively spliced MEFV transcripts in human synovial fibroblasts: a prominent splice

isoform lacks the C-terminal domain that is highly mutated in familial Mediterranean fever. Arthritis Rheum 2004; 50: 3679-3689.

Dinarello CA. Interleukin-1 beta, interleukin-18, and the interleukin-1 beta converting enzyme. Ann N Y Acad Sci 1998; 856: 1-11.

Dinarello CA. Blocking IL-1 in systemic inflammation. J Exp Med 2005a; 201: 1355-1359.

Dinarello CA. Interleukin-1beta. Crit Care Med 2005b; 33: S460-S462.

Dinarello CA. Mutations in cryopyrin: bypassing roadblocks in the caspase 1 inflammasome for interleukin-1beta secretion and disease activity. Arthritis Rheum 2007; 56: 2817-2822.

Dodé C, Le Dû N, Cuisset L, Letourneur F, Berthelot JM, Vaudour G, Meyrier A, Watts RA, Scott DG, Nicholls A, Granel B, Frances C, Garcier F, Edery P, Boulinguez S, Domergues JP, Delpech M, Grateau G. New mutations of CIAS1 that are responsible for Muckle-Wells syndrome and familial cold urticaria: a novel mutation underlies both syndromes. Am J Hum Genet 2002; 70: 1498-1506.

Dowds TA, Masumoto J, Chen FF, Ogura Y, Inohara N, Núñez G. Regulation of cryopyrin/Pypaf1 signaling by pyrin, the familial Mediterranean fever gene product. Biochem Biophys Res Commun 2003; 302: 575-580.

Drenth JP, van der Meer JW. Hereditary periodic fever. N Engl J Med 2001; 345: 1748-1757.

Drenth JP, van der Meer JW. The inflammasome--a linebacker of innate defense. N Engl J Med 2006; 355: 730-732.

Drossman DA. The functional gastrointestinal disorders and the Rome III process Gastroenterology 2006; 130: 1377-1390

Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum Hered 2008; 66: 87-98.

Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006; 314: 1461-1463.

Dunlop SP, Jenkins D, Neal KR, Spiller RC. Relative importance of enterochromaffin cell hyperplasia, anxiety, and depression in postinfectious IBS. Gastroenterology 2003a; 125: 1651-1659.

Dunlop SP, Jenkins D, Spiller RC. Distinctive clinical, psychological, and histological features of postinfective irritable bowel syndrome. Am J Gastroenterol 2003; 98:1578-1583.

Dunlop SP, Coleman NS, Blackshaw E, Perkins AC, Singh G, Marsden CA, Spiller RC. Abnormalities of 5-hydroxytryptamine metabolism in irritable bowel syndrome. Clin Gastroenterol Hepatol 2005; 3:349-357.

Dunlop SP, Hebden J, Campbell E, Naesdal J, Olbe L, Perkins AC, Spiller RC. Abnormal intestinal permeability in subgroups of diarrhea-predominant irritable bowel syndromes. Am J Gastroenterol 2006; 101:1288-1294.

DuPont AW. Postinfectious irritable bowel syndrome. Clin Infect Dis 2008; 46: 594-599.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. Science. 2005; 308: 1635-1638.

Eckle T, Füllbier L, Wehrmann M, Khoury J, Mittelbronn M, Ibla J, Rosenberger P, Eltzschig HK. Identification of ectonucleotidases CD39 and CD73 in innate protection during acute lung injury. J Immunol 2007; 178: 8127-8137.

Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JP. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. Am J Gastroenterol 2004; 99: 2393-2404.

Economou M, Pappas G. New global map of Crohn's disease: Genetic, environmental, and socioeconomic correlations. Inflamm Bowel Dis 2008; 14: 709-720.

Ellson CD, Davidson K, Ferguson GJ, O'Connor R, Stephens LR, Hawkins PT. Neutrophils from p40phox-/- mice exhibit severe defects in NADPH oxidase regulation and oxidant-dependent bacterial killing. J Exp Med 2006; 203: 1927-1937.

Elphick DA, Mahida YR. Paneth cells: their role in innate immunity and inflammatory disease. Gut 2005; 54: 1802-1809.

Elson CO, Cong Y, Weaver CT, Schoeb TR, McClanahan TK, Fick RB, Kastelein RA. Monoclonal anti-interleukin 23 reverses active colitis in a T cell-mediated model in mice. Gastroenterology 2007; 132: 2359-2370.

Ewaschuk JB, Backer JL, Churchill TA, Obermeier F, Krause DO, Madsen KL. Surface expression of Toll-like receptor 9 is upregulated on intestinal epithelial cells in response to pathogenic bacterial DNA. Infect Immun 2007; 75:2572-2579.

Fabry W .Wilhelm Fabry (1560-1624)--the other Fabricius. JAMA 1964; 190: 933.

Falk RH, Comenzo RL, Skinner M. The systemic amyloidoses. N Engl J Med 1997; 337: 898-909.

Farrokhyar F, Swarbrick ET, Irvine EJ. A critical review of epidemiological studies in inflammatory bowel disease. Scand J Gastroenterol 2001; 36:2-15.

Fedorak RN, Gangl A, Elson CO, Rutgeerts P, Schreiber S, Wild G, Hanauer SB, Kilian A, Cohard M, LeBeaut A, Feagan B. Recombinant human interleukin 10 in the treatment of patients with mild to moderately active Crohn's disease. The Interleukin 10 Inflammatory Bowel Disease Cooperative Study Group. Gastroenterology 2000; 119: 1473–1482.

Fedorak RN, Dieleman LA. Probiotics in the treatment of human inflammatory bowel diseases: update 2008. J Clin Gastroenterol 2008; 42: S97-103.

Feldmann J, Prieur AM, Quartier P, Berquin P, Certain S, Cortis E, Teillac-Hamel D, Fischer A, de Saint Basile G. Chronic infantile neurological cutaneous and articular syndrome is caused by mutations in CIAS1, a gene highly expressed in polymorphonuclear cells and chondrocytes. Am J Hum Genet 2002; 71: 198-203.

Feldmeyer L, Keller M, Niklaus G, Hohl D, Werner S, Beer HD. The inflammasome mediates UVB-induced activation and secretion of interleukin-1beta by keratinocytes. Curr Biol. 2007; 17: 1140-1145.

Ferguson DR, Kennedy I, Burton TJ. ATP is released from rabbit urinary bladder epithelial cells by hydrostatic pressure changes--a possible sensory mechanism? J Physiol 1997; 505: 503-511.

Fernandes-Alnemri T, Wu J, Yu JW, Datta P, Miller B, Jankowski W, Rosenberg S, Zhang J, Alnemri ES. The pyroptosome: a supramolecular assembly of ASC dimers mediating inflammatory cell death via caspase-1 activation. Cell Death Differ 2007; 14: 1590-1604.

Ferrari D, Pizzirani C, Adinolfi E, Lemoli RM, Curti A, Idzko M, Panther E, Di Virgilio F. The P2X7 receptor: a key player in IL-1 processing and release. J Immunol 2006; 176: 3877-3883.

Ferrero-Miliani L, Nielsen OH, Andersen PS, Girardin SE. Chronic inflammation: importance of NOD2 and NALP3 in interleukin-1beta generation. Clin Exp Immunol 2007; 147: 227-235.

Ferrier L, Mazelin L, Cenac N, Desreumaux P, Janin A, Emilie D, Colombel JF, Garcia-Villar R, Fioramonti J, Bueno L. Stress-induced disruption of colonic epithelial barrier: role of interferon-gamma and myosin light chain kinase in mice. Gastroenterology 2003; 125: 795-804.

Fesik SW. Insights into programmed cell death through structural biology. Cell 2000; 103: 273-282.

Fidder HH, Chowers Y, Lidar M, Sternberg M, Langevitz P, Livneh A. Crohn disease in patients with familial Mediterranean fever. Medicine (Baltimore) 2002; 81: 411-416.

Fidder H, Chowers Y, Ackerman Z, Pollak RD, Crusius JB, Livneh A, , Bar-Meir S, Avidan B, Shinhar Y. The familial Mediterranean fever (MEVF) gene as a modifier of Crohn's disease. Am J Gastroenterol 2005; 100: 338-343.

Fina D, Sarra M, Fantini MC, Rizzo A, Caruso R, Caprioli F, Stolfi C. Regulation of gut inflammation and th17 cell response by interleukin-21. Gastroenterology 2008; 134: 1038-1048.

Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, Lees CW, Onnie CM, Hanson CE, Blaszczyk K, Inouye M, Ewels P, Ravindrarajah R, Keniry A, Hunt S, Carter M, Watkins N, Ouwehand W, Lewis CM, Cardon L; Wellcome Trust Case Control Consortium, Lobo A, Forbes A, Sanderson J, Jewell DP, Mansfield JC, Deloukas P, Mathew CG, Parkes M, Satsangi J. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet 2008; 40: 710-712.

Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, Woo P. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. J Clin Invest 1998; 102:1369-1376.

Franchi L, Kanneganti TD, Dubyak GR, Núñez G. Differential requirement of P2X7 receptor and intracellular K+ for caspase-1 activation induced by intracellular and extracellular bacteria. J Biol Chem 2007; 282: 18810-18818.

Franchi L, Warner N, Viani K, Nuñez G. Function of Nod-like receptors in microbial recognition and host defense. Immunol Rev 2009; 227: 106-128.

Frank CF, Hostetter MK. Cleavage of E-cadherin: a mechanism for disruption of the intestinal epithelial barrier by Candida albicans. Transl Res 2007; 149:211-222.

Franke A, Hampe J, Rosenstiel P, Becker C, Wagner F, Häsler R, Little RD, Huse K, Ruether A, Balschun T, Wittig M, Elsharawy A, Mayr G, Albrecht M, Prescott NJ, Onnie CM, Fournier H, Keith T, Radelof U, Platzer M, Mathew CG, Stoll M, Krawczak M, Nürnberg P, Schreiber S. Systematic association mapping identifies NELL1 as a novel IBD disease gene. PLoS ONE 2007; 2:e691.

Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, Domingues FS, Albrecht M, Nothnagel M, Ellinghaus D, Sina C, Onnie CM, Weersma RK, Stokkers PC, Wijmenga C, Gazouli M, Strachan D, McArdle WL, Vermeire S, Rutgeerts P, Rosenstiel P, Krawczak M, Vatn MH; IBSEN study group, Mathew CG, Schreiber S. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet 2008a; 40: 1319-1323.

Franke A, Balschun T, Karlsen TH, Hedderich J, May S, Lu T, Schuldt D, Nikolaus S, Rosenstiel P, Krawczak M, Schreiber S. Replication of signals

from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. Nat Genet 2008b; 40: 713-715.

Franchimont D, Vermeire S, El Housni H, Pierik M, Van Steen K, Gustot T, Quertinmont E, Abramowicz M, Van Gossum A, Devière J, Rutgeerts P. 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: 987-992.

French FMF Consortium. A candidate gene for familial Mediterranean fever. Nat Genet 1997; 17: 25-31.

Friedrichs F, Stoll M. Role of discs large homolog 5. World J Gastroenterol 2006; 12: 3651-3656.

Fritz JH, Ferrero RL, Philpott DJ, Girardin SE. Nod-like proteins in immunity, inflammation and disease. Nat Immunol 2006; 7: 1250-1257.

Fujino S, Andoh A, Bamba S, Ogawa A, Hata K, Araki Y, Bamba T, Fujiyama Y. Increased expression of interleukin 17 in inflammatory bowel disease. Gut 2003; 52: 65-70.

Garg AX, Moist L, Matsell D, Thiessen-Philbrook HR, Haynes RB, Suri RS, Salvadori M, Ray J, Clark WF; Walkerton Health Study Investigators. Risk of hypertension and reduced kidney function after acute gastroenteritis from bacteria-contaminated drinking water. CMAJ 2005a; 173: 261-268.

Garg AX, Macnab J, Clark W, Ray JG, Marshall JK, Suri RS, Devereaux PJ, Haynes B; Walkerton Health Study Investigators. Long-term health sequelae following E. coli and campylobacter contamination of municipal

water. Population sampling and assessing non-participation biases. Can J Public Health 2005b; 96:125-130.

Garg AX, Marshall J, Salvadori M, Thiessen-Philbrook HR, Macnab J, Suri RS, Haynes RB, Pope J, Clark W; Walkerton Health Study Investigators. A gradient of acute gastroenteritis was characterized, to assess risk of long-term health sequelae after drinking bacterial-contaminated water. J Clin Epidemiol 2006; 59:421-428.

Gaya DR, Russell RK, Nimmo ER, Satsangi J. New genes in inflammatory bowel disease: lessons for complex diseases? Lancet 2006; 367: 1271-1284.

Gershon MD. Review article: roles played by 5-hydroxytryptamine in the physiology of the bowel. Aliment Pharmacol Ther 1999; 13 (Suppl 2):15-30.

Giaglis S, Mimidis K, Papadopoulos V, Thomopoulos K, Sidiropoulos P, Rafail S, Nikolopoulou V, Fragouli E, Kartalis G, Tzioufas A, Boumpas D, Ritis K. Increased frequency of mutations in the gene responsible for familial Mediterranean fever (MEFV) in a cohort of patients with ulcerative colitis: evidence for a potential disease-modifying effect? Dig Dis Sci 2006; 51: 687-692.

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. Metagenomic analysis of the human distal gut microbiome. Science. 2006; 312: 1355-1359.

Girardin SE, Tournebize R, Mavris M, Page AL, Li X, Stark GR, Bertin J, DiStefano PS, Yaniv M, Sansonetti PJ, Philpott DJ. CARD4/Nod1

mediates NF-kappaB and JNK activation by invasive Shigella flexneri. EMBO Rep 2001; 2: 736-742.

Girardin SE, Travassos LH, Hervé M, Blanot D, Boneca IG, Philpott DJ, Sansonetti PJ, Mengin-Lecreulx D. Peptidoglycan molecular requirements allowing detection by Nod1 and Nod2. J Biol Chem 2003; 278: 41702-41708.

Goldbach-Mansky R, Dailey NJ, Canna SW, Gelabert A, Jones J, Rubin BI, Kim HJ, Brewer C, Zalewski C, Wiggs E, Hill S, Turner ML, Karp BI, Aksentijevich I, Pucino F, Penzak SR, Haverkamp MH, Stein L, Adams BS, Moore TL, Fuhlbrigge RC, Shaham B, Jarvis JN, O'Neil K, Vehe RK, Beitz LO, Gardner G, Hannan WP, Warren RW, Horn W, Cole JL, Paul SM, Hawkins PN, Pham TH, Snyder C, Wesley RA, Hoffmann SC, Holland SM, Butman JA, Kastner DL. Neonatal-onset multisystem inflammatory disease responsive to interleukin-1beta inhibition. N Engl J Med 2006; 355: 581-592.

Goldfinger SE. Colchicine for familial Mediterranean fever. N Engl J Med 1972; 287:1302.

Gonsalkorale WM, Perrey C, Pravica V, Whorwell PJ, Hutchinson IV. Interleukin 10 genotypes in irritable bowel syndrome: evidence for an inflammatory component? Gut 2003; 52: 91-93.

Göring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet 2001; 69: 1357-1369.

Grabbe S, Schwarz T. Immunoregulatory mechanisms involved in elicitation of allergic contact hypersensitivity. Immunol Today 1998; 19: 37-44.

Granel B, Serratrice J, Disdier P, Weiller PJ. Dramatic improvement with anakinra in a case of chronic infantile neurological cutaneous and articular (CINCA) syndrome. Rheumatology (Oxford) 2005; 44: 689-690.

Grant SF, Hakonarson H. Microarray technology and applications in the arena of genome-wide association. Clin Chem 2008; 54: 1116-1124.

Greenberg BD, Tolliver TJ, Huang SJ, Li Q, Bengel D, Murphy DL. Genetic variation in the serotonin transporter promoter region affects serotonin uptake in human blood platelets. Am J Med Genet 1999; 88: 83-87.

Grierson JP, Meldolesi J. Shear stress-induced [Ca2+]i transients and oscillations in mouse fibroblasts are mediated by endogenously released ATP. J Biol Chem 1995; 270: 4451-4456.

Grütter C, Briand C, Capitani G, Mittl PR, Papin S, Tschopp J, Grütter MG. Structure of the PRYSPRY-domain: implications for autoinflammatory diseases. FEBS Lett 2006; 580: 99-106.

Grygorczyk R, Hanrahan JW. CFTR-independent ATP release from epithelial cells triggered by mechanical stimuli. Am J Physiol 1997; 272: C1058-1066.

Guarner F, Malagelada JR. Gut flora in health and disease. Lancet 2003; 361: 512-519.

Guilarte M, Santos J, de Torres I, Alonso C, Vicario M, Ramos L, Martínez C, Casellas F, Saperas E, Malagelada JR. Diarrhoea-predominant IBS

patients show mast cell activation and hyperplasia in the jejunum. Gut 2007; 56: 203-209.

Gurcel L, Abrami L, Girardin S, Tschopp J, van der Goot FG. Caspase-1 activation of lipid metabolic pathways in response to bacterial pore-forming toxins promotes cell survival. Cell 2006; 126: 1135-1145.

Guslandi M, Mezzi G, Sorghi M, Testoni PA. Saccharomyces boulardii in maintenance treatment of Crohn's disease. Dig Dis Sci 2000; 45: 1462–1464.

Gwee KA, Leong YL, Graham C, McKendrick MW, Collins SM, Walters SJ, Underwood JE, Read NW. The role of psychological and biological factors in postinfective gut dysfunction. Gut 1999; 44:400-406.

Gwee KA, Collins SM, Read NW, Rajnakova A, Deng Y, Graham JC, McKendrick MW,Moochhala SM. Increased rectal mucosal expression of interleukin 1beta in recently acquired post-infectious irritable bowel syndrome. Gut 2003; 52: 523-526.

Halvorson HA, Schlett CD, Riddle MS. Postinfectious irritable bowel syndrome--a meta-analysis. Am J Gastroenterol 2006; 101: 1894-1899.

Hampe J, Heymann K, Kruis W, Raedler A, Fölsch UR, Schreiber S. Anticipation in inflammatory bowel disease: a phenomenon caused by an accumulation of confounders. Am J Med Genet 2000; 92: 178-183.

Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega FM, Briggs J, Günther S, Prescott NJ, Onnie CM, Häsler R, Sipos B, Fölsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S. A genome-wide association scan of nonsynonymous

SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 2007; 39: 207-211.

Harton JA, Linhoff MW, Zhang J, Ting JP. Cutting edge: CATERPILLER: a large family of mammalian genes containing CARD, pyrin, nucleotide-binding, and leucine-rich repeat domains. J Immunol 2002; 169: 4088-4093.

Hawkes CH. Twin studies in medicine--what do they tell us? QJM 1997; 90: 311-321.

Hawkins PN, Lachmann HJ, McDermott MF. Interleukin-1-receptor antagonist in the Muckle-Wells syndrome. N Engl J Med 2003; 348: 2583-2584.

Hawkins PN, Lachmann HJ, Aganna E, McDermott MF. Spectrum of clinical features in Muckle-Wells syndrome and response to anakinra. Arthritis Rheum 2004; 50: 607-612.

Hazama A, Shimizu T, Ando-Akatsuka Y, Hayashi S, Tanaka S, Maeno E, Okada Y. Swelling-induced, CFTR-independent ATP release from a human epithelial cell line: lack of correlation with volume-sensitive cl(-) channels. J Gen Physiol 1999; 114: 525-533.

Heesen M, Bloemeke B, Heussen N, Kunz D. Can the interleukin-6 response to endotoxin be predicted? Studies of the influence of a promoter polymorphism of the interleukin-6 gene, gender, the density of the endotoxin receptor CD14, and inflammatory cytokines. Crit Care Med 2002; 30:664-669.

Heils A, Teufel A, Petri S, Stöber G, Riederer P, Bengel D, Lesch KP. Allelic variation of human serotonin transporter gene expression. J Neurochem 1996; 66: 2621-2624.

Hentgen V, Despert V, Leprêtre AC, Cuisset L, Chevrant-Breton J, Jégo P, Chalès G, Gall EL, Delpech M, Grateau G. Intrafamilial variable phenotypic expression of a CIAS1 mutation: from Muckle-Wells to chronic infantile neurological cutaneous and articular syndrome. J Rheumatol 2005; 32: 747-751.

Heresbach D, Gulwani-Akolkar B, Lesser M, Akolkar PN, Lin XY, Heresbach-Le Berre N, Bretagne JF, Katz S, Silver J. Anticipation in Crohn's disease may be influenced by gender and ethnicity of the transmitting parent. Am J Gastroenterol 1998; 93: 2368-2372.

Hirano Y, Hendil KB, Yashiroda H, Iemura S, Nagane R, Hioki Y, Natsume T, Tanaka K, Murata S. A heterodimeric complex that promotes the assembly of mammalian 20S proteasomes. Nature 2005; 437: 1381-1385.

Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005; 6: 95-108.

Hisamatsu T, Suzuki M, Reinecker HC, Nadeau WJ, McCormick BA, Podolsky DK. CARD15/NOD2 functions as an antibacterial factor in human intestinal epithelial cells. Gastroenterology 2003; 124: 993-1000.

Hoffman HM, Mueller JL, Broide DH, Wanderer AA, Kolodner RD. Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome. Nat Genet 2001a; 29: 301-305.

Hoffman HM, Wanderer AA, Broide DH. Familial cold autoinflammatory syndrome: phenotype and genotype of an autosomal dominant periodic fever. J Allergy Clin Immunol 2001b; 108: 615-620.

Hoffman H, Gregory S, Mueller J, Tresierras M, Broide D, Wanderer A, Kolodner R. Fine structure mapping of CIAS1: Identification of an ancestral haplotype and a common FCAS mutation L353P. Human Genetics 2003; 112: 209-216.

Hoffman HM, Rosengren S, Boyle DL, Cho JY, Nayar J, Mueller JL, Anderson JP, Wanderer AA, Firestein GS. Prevention of cold-associated acute inflammation in familial cold autoinflammatory syndrome by interleukin-1 receptor antagonist. Lancet 2004; 364: 1779-1785.

Hogquist KA, Nett MA, Unanue ER, Chaplin DD. Interleukin 1 is processed and released during apoptosis. Proc Natl Acad Sci USA 1991; 88: 8485-8489.

Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5′→3′ exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci USA 1991; 88: 7276–7280.

Holtkamp W, Stollberg T, Reis HE. Serum interleukin-6 is related to disease activity but not disease specificity in inflammatory bowel disease. J Clin Gastroenterol 1995; 20:123-6.

Hölttä V, Klemetti P, Sipponen T, Westerholm-Ormio M, Kociubinski G, Salo H, Räsänen L, Kolho KL, Färkkilä M, Savilahti E, Vaarala O. IL-23/IL-17 immunity as a hallmark of Crohn's disease. Inflamm Bowel Dis 2008; 14: 1175-1184.

Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI. Molecular analysis of commensal host-microbial relationships in the intestine. Science. 2001; 291: 881-884.

Hsu TL, Chang YC, Chen SJ, Liu YJ, Chiu AW, Chio CC, Chen L, Hsieh SL. Modulation of dendritic cell differentiation and maturation by decoy receptor 3. J Immunol 2002; 168: 4846-4853.

Hsu TM, Chen X, Duan S, Miller RD, Kwok PY. Universal SNP genotyping assay with fluorescence polarization detection. Biotechniques 2001; 31: 560, 562, 564-568.

Hue S, Ahern P, Buonocore S, Kullberg MC, Cua DJ, McKenzie BS, Powrie F, Maloy KJ. Interleukin-23 drives innate and T cell-mediated intestinal inflammation. J Exp Med 2006; 203: 2473-2483.

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001; 411: 599-603.

Hugot JP, Zaccaria I, Cavanaugh J, Yang H, Vermeire S, Lappalainen M, Schreiber S, Annese V, Jewell DP, Fowler EV, Brant SR, Silverberg MS, Cho J, Rioux JD, Satsangi J, Parkes M; for the IBD International Genetics Consortium. Prevalence of CARD15/NOD2 mutations in Caucasian healthy people. Am J Gastroenterol 2007; 102: 1259-1267.

Hulisz D. The burden of illness of irritable bowel syndrome: current challenges and hope for the future. J Manag Care Pharm 2004; 10:299-309.

Ilnyckyj A, Balachandra B, Elliott L, Choudhri S, Duerksen DR. Post-traveler's diarrhea irritable bowel syndrome: a prospective study. Am J Gastroenterol 2003; 98: 596-599.

Inohara N, Ogura Y, Fontalba A, Gutierrez O, Pons F, Crespo J, Fukase K, Inamura S, Kusumoto S, Hashimoto M, Foster SJ, Moran AP, Fernandez-Luna JL, Nuñez G. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. J Biol Chem 2003; 278: 5509-5512.

International HapMap Consortium. A haplotype map of the human genome. Nature 2005; 437: 1299-1320.

International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007; 449: 851-861.

Izcue A, Hue S, Buonocore S, Arancibia-Cárcamo CV, Ahern PP, Iwakura Y, Maloy KJ, Powrie F. Interleukin-23 restrains regulatory T cell activity to drive T cell-dependent colitis. Immunity 2008; 28: 559-570.

Janeway CA Jr, Medzhitov R. Innate immune recognition. Annu Rev Immunol 2002; 20: 197-216.

Jéru I, Papin S, L'hoste S, Duquesnoy P, Cazeneuve C, Camonis J, Amselem S. Interaction of pyrin with 14.3.3 in an isoform-specific and phosphorylation-dependent manner regulates its translocation to the nucleus. Arthritis Rheum 2005; 52: 1848-1857.

Jin Y, Mailloux CM, Gowan K, Riccardi SL, LaBerge G, Bennett DC, Fain PR, Spritz RA. NALP1 in vitiligo-associated multiple autoimmune disease. N Engl J Med. 2007; 356: 1216-1225.

Joazeiro CA, Weissman AM. RING finger proteins: mediators of ubiquitin ligase activity. Cell 2000; 102: 549-552.

Johnstone RF, Dolen WK, Hoffman HM. A large kindred with familial cold autoinflammatory syndrome. Ann Allergy Asthma Immunol 2003; 90: 233-237.

Kalantar JS, Locke GR 3rd, Zinsmeister AR, Beighley CM, Talley NJ. Familial aggregation of irritable bowel syndrome: a prospective study. Gut 2003; 52: 1703-1707.

Kanazawa M, Endo Y, Whitehead WE, Kano M, Hongo M, Fukudo S. Patients and nonconsulters with irritable bowel syndrome reporting a parental history of bowel problems have more impaired psychological distress. Dig Dis Sci. 2004; 49: 1046-1053.

Kanazawa N, Okafuji I, Kambe N, Nishikomori R, Nakata-Hizume M, Nagai S, Fuji A, Yuasa T, Manki A, Sakurai Y, Nakajima M, Kobayashi H, Fujiwara I, Tsutsumi H, Utani A, Nishigori C, Heike T, Nakahata T, Miyachi Y. Early-onset sarcoidosis and CARD15 mutations with constitutive nuclear factor-kappaB activation: common genetic etiology with Blau syndrome. Blood 2005; 105: 1195-1197.

Kanneganti TD, Ozören N, Body-Malapel M, Amer A, Park JH, Franchi L, Whitfield J, Barchet W, Colonna M, Vandenabeele P, Bertin J, Coyle A, Grant EP, Akira S, Núñez G. Bacterial RNA and small antiviral

compounds activate caspase-1 through cryopyrin/Nalp3. Nature 2006a; 440: 233-236.

Kanneganti TD, Body-Malapel M, Amer A, Park JH, Whitfield J, Franchi L, Taraporewala ZF, Miller D, Patton JT, Inohara N, Núñez G. Critical role for Cryopyrin/Nalp3 in activation of caspase-1 in response to viral infection and double-stranded RNA. J Biol Chem 2006b; 281: 36560-36568.

Kanneganti TD, Lamkanfi M, Kim YG, Chen G, Park JH, Franchi L, Vandenabeele P, Núñez G. Pannexin-1-mediated recognition of bacterial molecules activates the cryopyrin inflammasome independent of Toll-like receptor signaling. Immunity 2007; 26: 433-443.

Karban A, Dagan E, Eliakim R, Herman A, Nesher S, Weiss B, Berkowitz D, Shamir R, Gershoni-Baruch R. Prevalence and significance of mutations in the familial Mediterranean fever gene in patients with Crohn's disease. Genes Immun 2005; 6: 134-139.

Karin M, Lawrence T, Nizet V. Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer. Cell 2006; 124: 823-835.

Kastner DL, Aksentijevich I. Intermittent and periodic arthritis syndromes. In: Koopman WJ, Moreland LW, eds. Arthritis and Allied Conditions (15th ed). Philadelphia: Lippincott Williams and Wilkins; 2005:1411-1461.

Kawai T, Akira S. TLR signaling. Semin Immunol. 2007; 19: 24-32.

Kelsall B. Getting to the guts of NOD2. Nat Med 2005; 11: 383-384.

Kile RL, Rusk HA. A case of cold urticaria with an unusual family history. JAMA 1940; 114: 1067–1068

Kim HJ, Camilleri M, Carlson PJ, Cremonini F, Ferber I, Stephens D, McKinzie S, Zinsmeister AR, Urrutia R. Association of distinct alpha(2) adrenoceptor and serotonin transporter polymorphisms with constipation and somatic symptoms in functional gastrointestinal disorders. Gut 2004; 53: 829-837.

Kobayashi KS, Chamaillard M, Ogura Y, Henegariu O, Inohara N, Nuñez G, Flavell RA. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. Science 2005; 307: 731-734.

Kostura MJ, Tocci MJ, Limjuco G, Chin J, Cameron P, Hillman AG, Chartrain NA, Schmidt JA. Identification of a monocyte specific pre-interleukin 1 beta convertase activity. Proc Natl Acad Sci USA 1989; 86: 5227-5231.

Kozuch PL, Hanauer SB. Treatment of inflammatory bowel disease: a review of medical therapy. World J Gastroenterol 2008; 14: 354-377.

Kramer M, Netea MG, de Jong DJ, Kullberg BJ, Adema GJ. Impaired dendritic cell function in Crohn's disease patients with NOD2 3020insC mutation. J Leukoc Biol 2006; 79: 860-866.

Krieg AM. CpG motifs in bacterial DNA and their immune effects. Annu Rev Immunol 2002; 20: 709-760.

Kuballa P, Huett A, Rioux JD, Daly MJ, Xavier RJ. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated ATG16L1 variant. PLoS ONE 2008; 3: e3391.

Kuehn BM. 1000 Genomes Project promises closer look at variation in human genome. JAMA 2008; 300: 2715.

Kugathasan S, Loizides A, Babusukumar U, McGuire E, Wang T, Hooper P, Nebel J, Kofman G, Noel R, Broeckel U, Tolia V. Comparative phenotypic and CARD15 mutational analysis among African American, Hispanic, and White children with Crohn's disease. Inflamm Bowel Dis 2005; 11: 631-638.

Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PM, Imielinski M, Guthery SL, Cucchiara S, Kim CE, Frackelton EC, Annaiah K, Glessner JT, Santa E, Willson T, Eckert AW, Bonkowski E, Shaner JL, Smith RM, Otieno FG, Peterson N, Abrams DJ, Chiavacci RM, Grundmeier R, Mamula P, Tomer G, Piccoli DA, Monos DS, Annese V, Denson LA, Grant SF, Hakonarson H. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. Nat Genet 2008; 40: 1211-1215.

Kuida K, Lippke JA, Ku G, Harding MW, Livingston DJ, Su MS, Flavell RA. Altered cytokine export and apoptosis in mice deficient in interleukin-1 beta converting enzyme. Science 1995; 267: 2000-2003.

Kullberg MC, Jankovic D, Feng CG, Hue S, Gorelick PL, McKenzie BS, Cua DJ, Powrie F, Cheever AW, Maloy KJ, Sher A. IL-23 plays a key role in Helicobacter hepaticus-induced T cell-dependent colitis. J Exp Med 2006; 203: 2485-2494.

Kummer JA, Broekhuizen R, Everett H, Agostini L, Kuijk L, Martinon F, van Bruggen R, Tschopp J. Inflammasome components NALP 1 and 3 show distinct but separate expression profiles in human tissues suggesting a site-specific role in the inflammatory response. J Histochem Cytochem 2007; 55: 443-452.

Kurata JH, Kantor-Fish S, Frankl H, Godby P, Vadheim CM. Crohn's disease among ethnic groups in a large health maintenance organization. Gastroenterology 1992; 102: 1940-1948.

La JH, Kim TW, Sung TS, Kim HJ, Kim JY, Yang IS. Role of mucosal mast cells in visceral hypersensitivity in a rat model of irritable bowel syndrome. J Vet Sci 2004; 5: 319-324.

Lazarus R, Klimecki WT, Raby BA, Vercelli D, Palmer LJ, Kwiatkowski DJ, Silverman EK, Martinez F, Weiss ST. Single-nucleotide polymorphisms in the Toll-like receptor 9 gene (TLR9): frequencies, pairwise linkage disequilibrium, and haplotypes in three U.S. ethnic groups and exploratory case-control disease association studies. Genomics 2003; 81:85-91.

Lembo A, Zaman M, Jones M, Talley NJ. Influence of genetics on irritable bowel syndrome, gastro-oesophageal reflux and dyspepsia: a twin study. Aliment Pharmacol Ther 2007; 25: 1343-1350.

Lennard-Jones JE. Classification of inflammatory bowel disease. Scan J Gastroenterol 1989; Suppl 170: 2-6.

Leong RW, Armuzzi A, Ahmad T, Wong ML, Tse P, Jewell DP, Sung JJ. NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. Aliment Pharmacol Ther 2003; 17: 1465-1470.

Leong SA, Barghout V, Birnbaum HG, Thibeault CE, Ben-Hamadi R, Frech F, Ofman JJ. The economic consequences of irritable bowel syndrome: a US employer perspective. Arch Intern Med 2003; 163: 929-935.

Levy RL, Whitehead WE, Von Korff MR, Feld AD. Intergenerational transmission of gastrointestinal illness behavior. Am J Gastroenterol 2000; 95: 451-456.

Levy RL, Jones KR, Whitehead WE, Feld SI, Talley NJ, Corey LA. Irritable bowel syndrome in twins: heredity and social learning both contribute to etiology. Gastroenterology 2001; 121: 799-804.

Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 2006; 124: 837-848.

Li J, Moran T, Swanson E, Julian C, Harris J, Bonen DK, Hedl M, Nicolae DL, Abraham C, Cho JH. Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. Hum Mol Genet 2004; 13: 1715-1725.

Li LC, Chui RM, Sasaki M, Nakajima K, Perinchery G, Au HC, Nojima D, Carroll P, Dahiya R. A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities. Cancer Res 2000 15; 60:873-6.

Li P, Allen H, Banerjee S, Franklin S, Herzog L, Johnston C, McDowell J, Paskind M, Rodman L, Salfeld J, Towne E, Tracey D, Wardwell S, Wei FY, Wong W, Kamen R, Seshadri T. Mice deficient in IL-1 beta-converting enzyme are defective in production of mature IL-1 beta and resistant to endotoxic shock. Cell 1995; 80: 401-11.

Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A, Demarche B, Gut I, Heath S, Foglio M, Liang L, Laukens D, Mni M, Zelenika D, Van Gossum A, Rutgeerts P,

Belaiche J, Lathrop M, Georges M. 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: e58.

Lichtenberger GS, Flavell RA, Alexopoulou L. Innate immunity and apoptosis in IBD. Inflamm Bowel Dis. 2004; 10 (Suppl 1): S58-S62.

Lidar M, Livneh A. Familial Mediterranean fever: clinical, molecular and management advancements. Neth J Med 2007; 65: 318-324.

Liebregts T, Adam B, Bertel A, Jones S, Schulze J, Enders C, Sonnenborn U, Lackner K, Holtmann G. Effect of E. coli Nissle 1917 on post-inflammatory visceral sensory function in a rat model. Neurogastroenterol Motil. 2005; 17: 410-414.

Liebregts T, Adam B, Bredack C, Röth A, Heinzel S, Lester S, Downie-Doyle S, Smith E, Drew P, Talley NJ, Holtmann G. Immune activation in patients with irritable bowel syndrome. Gastroenterology 2007; 132:913-920.

Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. Am J Hum Genet 2007; 80: 531-538.

Locke GR 3rd, Zinsmeister AR, Talley NJ, Fett SL, Melton LJ 3rd. Familial association in adults with functional gastrointestinal disorders. Mayo Clin Proc 2000; 75: 907-912.

Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. Gastroenterology 2004; 126: 1504-1517.

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003; 33: 177-182.

Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. Gastroenterology 2006; 130: 1480-1491.

Low JH, Williams FA, Yang X, Cullen S, Colley J, Ling KL, Armuzzi A, Ahmad T, Neville MJ, Dechairo BM, Walton R, Lench NJ, Jewell DP. Inflammatory bowel disease is linked to 19p13 and associated with ICAM-1. Inflamm Bowel Dis 2004; 10: 173-181.

MacDonald TT, Monteleone G. Immunity, inflammation, and allergy in the gut. Science 2005a; 307: 1920-1925.

MacDonald TT, Gordon JN. Bacterial regulation of intestinal immune responses. Gastroenterol Clin North Am 2005b; 34: 401-12, vii-viii.

Maeda S, Hsu LC, Liu H, Bankston LA, Iimura M, Kagnoff MF, Eckmann L, Karin M. Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. Science 2005; 307: 734-738.

Manji GA, Wang L, Geddes BJ, Brown M, Merriam S, Al-Garawi A, Mak S, Lora JM, Briskin M, Jurman M, Cao J, DiStefano PS, Bertin J. PYPAF1, a PYRIN-containing Apaf1-like protein that assembles with ASC and regulates activation of NF-kappa B. J Biol Chem. 2002; 277: 11570-11575.

Mannon PJ, Fuss IJ, Mayer L, Elson CO, Sandborn WJ, Present D, Dolin B, Goodman N, Groden C, Hornung RL, Quezado M, Yang Z, Neurath MF, Salfeld J, Veldman GM, Schwertschlag U, Strober W; Anti-IL-12 Crohn's

Disease Study Group. Anti-interleukin-12 antibody for active Crohn's disease. N Engl J Med 2004; 351: 2069-2079.

Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008; 118: 1590-1605.

Mansfield E, Chae JJ, Komarow HD, Brotz TM, Frucht DM, Aksentijevich I, Kastner DL. The familial Mediterranean fever protein, pyrin, associates with microtubules and colocalizes with actin filaments. Blood 2001; 98: 851-859.

Marangoni F, Trifari S, Scaramuzza S, Panaroni C, Martino S, Notarangelo LD, Baz Z, Metin A, Cattaneo F, Villa A, Aiuti A, Battaglia M, Roncarolo MG, Dupré L. WASP regulates suppressor activity of human and murine CD4(+)CD25(+)FOXP3(+) natural regulatory T cells. J Exp Med 2007; 204: 369-380.

Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet 2008; 24: 133-141.

Mariathasan S, Newton K, Monack DM, Vucic D, French DM, Lee WP, Roose-Girma M, Erickson S, Dixit VM. Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. Nature 2004; 430: 213-218.

Mariathasan S, Weiss DS, Newton K, McBride J, O'Rourke K, Roose-Girma M, Lee WP, Weinrauch Y, Monack DM, Dixit VM. Cryopyrin activates the inflammasome in response to toxins and ATP. Nature 2006; 440: 228-232.

Mariathasan S, Monack DM. Inflammasome adaptors and sensors: intracellular regulators of infection and inflammation. Nat Rev Immunol 2007; 7: 31-40.

Marshall JK, Thabane M, Garg AX, Clark W, Meddings J, Collins SM; WEL Investigators. Intestinal permeability in patients with irritable bowel syndrome after a waterborne outbreak of acute gastroenteritis in Walkerton, Ontario. Aliment Pharmacol Ther 2004; 20:1317-1322.

Marshall JK, Thabane M, Garg AX, Clark WF, Salvadori M, Collins SM; Walkerton Health Study Investigators. Incidence and epidemiology of irritable bowel syndrome after a large waterborne outbreak of bacterial dysentery. Gastroenterology 2006; 131: 445-450.

Marteau P, Lemann M, Seksik P, Laharie D, Colombel JF, Bouhnik Y, Cadiot G, Soulé JC, Bourreille A, Metman E, Lerebours E, Carbonnel F, Dupas JL, Veyrac M, Coffin B, Moreau J, Abitbol V, Blum-Sperisen S, Mary JY. Ineffectiveness of Lactobacillus johnsonii LA1 for prophylaxis of postoperative recurrence in Crohn's disease: a randomised, double-blind, placebo-controlled GETAID trial. Gut 2006; 55: 842-847.

Martinon F, Hofmann K, Tschopp J. The pyrin domain: a possible member of the death domain-fold family implicated in apoptosis and inflammation. Curr Biol 2001; 11: R118-120.

Martinon F, Burns K, Tschopp J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. Mol Cell 2002; 10: 417-426.

Martinon F, Agostini L, Meylan E, Tschopp J. Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome. Curr Biol 2004; 14: 1929-1934.

Martinon F, Tschopp J. NLRs join TLRs as innate sensors of pathogens. Trends Immunol 2005; 26: 447-454.

Martinon F, Pétrilli V, Mayor A, Tardivel A, Tschopp J. Gout-associated uric acid crystals activate the NALP3 inflammasome. Nature 2006a; 440: 237-241.

Martinon F, Glimcher LH. Gout: new insights into an old disease. J Clin Invest 2006b; 116: 2073-2075.

Martinon F, Tschopp J. Inflammatory caspases and inflammasomes: master switches of inflammation. Cell Death Differ 2007; 14: 10-22.

Martinon F. Detection of immune danger signals by NALP3. J Leukoc Biol 2008; 83: 507-511.

Masters SL, Lobito AA, Chae J, Kastner DL. Recent advances in the molecular pathogenesis of hereditary recurrent fevers. Curr Opin Allergy Clin Immunol 2006a; 6: 428-433.

Masters SL, Yao S, Willson TA, Zhang JG, Palmer KR, Smith BJ, Babon JJ, Nicola NA, Norton RS, Nicholson SE. The SPRY domain of SSB-2 adopts a novel fold that presents conserved Par-4-binding residues. Nat Struct Mol Biol 2006b; 13: 77-84.

Masumoto J, Taniguchi S, Ayukawa K, Sarvotham H, Kishino T, Niikawa N, Hidaka E, Katsuyama T, Higuchi T, Sagara J. ASC, a novel 22-kDa protein, aggregates during apoptosis of human promyelocytic leukemia HL-60 cells. J Biol Chem 1999; 274: 33835-3838.

Masumoto J, Dowds TA, Schaner P, Chen FF, Ogura Y, Li M, Zhu L, Katsuyama T, Sagara J, Taniguchi S, Gumucio DL, Núñez G, Inohara N. ASC is an activating adaptor for NF-kappa B and caspase-8-dependent apoptosis. Biochem Biophys Res Commun 2003; 303: 69-73.

Matsuda A, Suzuki Y, Honda G, Muramatsu S, Matsuzaki O, Nagano Y, Doi T, Shimotohno K, Harada T, Nishida E, Hayashi H, Sugano S. Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. Oncogene 2003; 22: 3307-3318.

Matzinger P. Tolerance, danger, and the extended family. Annu Rev Immunol 1994; 12: 991-1045.

Matzinger P. The danger model: a renewed sense of self. Science. 2002; 296: 301-305.

Maxion-Bergemann S, Thielecke F, Abel F, Bergemann R. Costs of irritable bowel syndrome in the UK and US. Pharmacoeconomics 2006; 24: 21-37.

Mayor A, Martinon F, De Smedt T, Pétrilli V, Tschopp J. A crucial function of SGT1 and HSP90 in inflammasome activity links mammalian and plant innate immune responses. Nat Immunol 2007; 8: 497-503.

McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, Zody MC, Hall JL, Brant SR, Cho JH, Duerr RH, Silverberg MS, Taylor KD, Rioux JD, Altshuler D, Daly MJ, Xavier RJ. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nat Genet 2008; 40: 1107-1112.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008; 9: 356-369.

McDermott MF. A common pathway in periodic fever syndromes. Trends Immunol 2004; 25:457-460.

McDermott MF, Tschopp J. From inflammasomes to fevers, crystals and hypertension: how basic research explains inflammatory diseases. Trends Mol Med 2007; 13: 381-388.

McGovern DP, Hysi P, Ahmad T, van Heel DA, Moffatt MF, Carey A, Cookson WO, Jewell DP. Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. Hum Mol Genet 2005; 14: 1245-1250.

McKendrick MW, Read NW. Irritable bowel syndrome--post salmonella infection. J Infect 1994; 29(1):1-3.

McPeek MS, Wu X, Ober C. Best linear unbiased allele-frequency estimation in complex pedigrees. Biometrics 2004; 60: 359-367.

Medzhitov R, Preston-Hurlburt P, Janeway CA Jr. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature 1997; 388: 394-397.

Medzhitov R, Janeway CA Jr. An ancient system of host defense. Curr Opin Immunol 1998;10 : 12-15.

Meylan E, Tschopp J, Karin M. Intracellular pattern recognition receptors in the host response. Nature 2006; 442: 39-44.

Miceli-Richard C, Lesage S, Rybojad M, Prieur AM, Manouvrier-Hanu S, Häfner R, Chamaillard M, Zouali H, Thomas G, Hugot JP. CARD15 mutations in Blau syndrome. Nat Genet 2001; 29: 19-20.

Milhavet F, Cuisset L, Hoffman HM, Slim R, El-Shanti H, Aksentijevich I, Lesage S, Waterham H, Wise C, Sarrauste de Menthiere C, Touitou I. The infevers autoinflammatory mutation online registry: update with new genes and functions. Hum Mutat. 2008; 29: 803-808.

Milner P, Bodin P, Loesch A, Burnstock G. Rapid release of endothelin and ATP from isolated aortic endothelial cells exposed to increased flow. Biochem Biophys Res Commun 1990; 170: 649-656.

Mohammed I, Cherkas LF, Riley SA, Spector TD, Trudgill NJ. Genetic influences in irritable bowel syndrome: a twin study. Am J Gastroenterol 2005; 100: 1340-1344.

Morris-Yates A, Talley NJ, Boyce PM, Nandurkar S, Andrews G. Evidence of a genetic contribution to functional bowel disorder. Am J Gastroenterol 1998; 93: 1311-1317.

Morrison AC, Wilson CB, Ray M, Correll PH. Macrophage-stimulating protein, the ligand for the stem cell-derived tyrosine kinase/RON receptor tyrosine kinase, inhibits IL-12 production by primary peritoneal macrophages stimulated with IFN-gamma and lipopolysaccharide. J Immunol 2004; 172: 1825-1832.

Muckle TJ, Wells M. Urticaria, deafness, and amyloidosis: a new heredo-familial syndrome. Q J Med 1962; 31: 235-248.

Münz C. Autophagy and antigen presentation. Cell Microbiol 2006; 8: 891-898.

Murakami M, Yoshihara K, Shimbara S, Lambeau G, Singer A, Gelb MH, Sawada M, Inagaki N, Nagai H, Kudo I. Arachidonate release and eicosanoid generation by group IIE phospholipase A(2). Biochem Biophys Res Commun 2002; 292: 689-696.

Murphy CA, Langrish CL, Chen Y, Blumenschein W, McClanahan T, Kastelein RA, Sedgwick JD, Cua DJ. Divergent pro- and antiinflammatory roles for IL-23 and IL-12 in joint autoimmune inflammation. J Exp Med 2003; 198: 1951-1957.

Nakagawa I, Amano A, Mizushima N, Yamamoto A, Yamaguchi H, Kamimoto T, Nara A, Funao J, Nakata M, Tsuda K, Hamada S, Yoshimori T. Autophagy defends cells against invading group A Streptococcus. Science 2004; 306: 1037-1040.

Nakamura M, Saito H, Kasanuki J, Tamura Y, Yoshida S. Cytokine production in patients with inflammatory bowel disease. Gut. 1992; 33:933-937.

Neal KR, Hebden J, Spiller R. Prevalence of gastrointestinal symptoms six months after bacterial gastroenteritis and risk factors for development of the irritable bowel syndrome: postal survey of patients. BMJ. 1997; 314:779-782.

Neal KR, Barker L, Spiller RC. Prognosis in post-infective irritable bowel syndrome: a six year follow up study. Gut 2002; 51: 410-413.

Neurath MF, Fuss I, Kelsall BL, Stüber E, Strober W. Antibodies to interleukin 12 abrogate established experimental colitis in mice. J Exp Med 1995; 182: 1281-1290.

Neurath MF. IL-23: a master regulator in Crohn disease. Nat Med 2007; 13: 26-28.

Neven B, Callebaut I, Prieur AM, Feldmann J, Bodemer C, Lepore L, Derfalvi B, Benjaponpitak S, Vesely R, Sauvain MJ, Oertle S, Allen R, Morgan G, Borkhardt A, Hill C, Gardner-Medwin J, Fischer A, de Saint Basile G. Molecular basis of the spectral expression of CIAS1 mutations associated with phagocytic cell-mediated autoinflammatory disorders CINCA/NOMID, MWS, and FCU. Blood 2004; 103: 2809-2815.

Neven B, Prieur AM, Quartier dit Maire P; Medscape. Cryopyrinopathies: update on pathogenesis and treatment. Nat Clin Pract Rheumatol 2008; 4: 481-489

Nguyen GC, Torres EA, Regueiro M, Bromfield G, Bitton A, Stempak J, Dassopoulos T, Schumm P, Gregory FJ, Griffiths AM, Hanauer SB, Hanson J, Harris ML, Kane SV, Orkwis HK, Lahaie R, Oliva-Hemker M, Pare P, Wild GE, Rioux JD, Yang H, Duerr RH, Cho JH, Steinhart AH, Brant SR, Silverberg MS. Inflammatory bowel disease characteristics among African Americans, Hispanics, and non-Hispanic Whites: characterization of a large North American cohort. Am J Gastroenterol 2006; 101: 1012-1023.

Niv Y, Abuksis G, Fraser GM. Epidemiology of ulcerative colitis in Israel: a survey of Israeli kibbutz settlements. Am J Gastroenterol 2000; 95: 693-698.

Nusrat A, Turner JR, Madara JL. Molecular physiology and pathophysiology of tight junctions. IV. Regulation of tight junctions by extracellular stimuli:

nutrients, cytokines, and immune cells. Am J Physiol Gastrointest Liver Physiol 2000; 279: G851-857.

Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nuñez G, Cho JH. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001; 411: 603-606.

Ogura Y, Sutterwala FS, Flavell RA. The inflammasome: first line of the immune response to cell stress. Cell 2006; 126: 659-662.

Ohman L, Simrén M. New insights into the pathogenesis and pathophysiology of irritable bowel syndrome. Dig Liver Dis 2007; 39: 201-215.

Okayasu I, Hatakeyama S, Yamada M, Ohkusa T, Inagaki Y, Nakaya R. A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. Gastroenterology 1990; 98: 694-702.

Okhuysen PC, Jiang ZD, Carlin L, Forbes C, DuPont HL. Post-diarrhea chronic intestinal symptoms and irritable bowel syndrome in NorthAmerican travelers to Mexico. Am J Gastroenterol 2004; 99: 1774-1778.

Oppmann B, Lesley R, Blom B, Timans JC, Xu Y, Hunte B, Vega F, Yu N, Wang J, Singh K, Zonin F, Vaisberg E, Churakova T, Liu M, Gorman D, Wagner J, Zurawski S, Liu Y, Abrams JS, Moore KW, Rennick D, de Waal-Malefyt R, Hannum C, Bazan JF, Kastelein RA. Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. Immunity 2000; 13: 715-725.

Orholm M, Munkholm P, Langholz E, Nielsen OH, Sørensen TI, Binder V. Familial occurrence of inflammatory bowel disease. N Engl J Med 1991; 324: 84-88.

Orholm M, Binder V, Sørensen TI, Rasmussen LP, Kyvik KO. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. Scand J Gastroenterol 2000; 35: 1075-1081.

O'Sullivan M, Clayton N, Breslin NP, Harman I, Bountra C, McLaren A, O'Morain CA. Increased mast cells in the irritable bowel syndrome. Neurogastroenterol Motil 2000; 12: 449-457.

Pabst O, Förster R, Lipp M, Engel H, Arnold HH. NKX2.3 is required for MAdCAM-1 expression and homing of lymphocytes in spleen and mucosa-associated lymphoid tissue. EMBO J 2000; 19: 2015-2023.

Packey CD, Sartor RB. Interplay of commensal and pathogenic bacteria, genetic mutations, and immunoregulatory defects in the pathogenesis of inflammatory bowel diseases. J Intern Med 2008; 263: 597-606.

Palm NW, Medzhitov R. Pattern recognition receptors and control of adaptive immunity. Immunol Rev 2009; 227:221-33.

Palmer C, Bik EM, Digiulio DB, Relman DA, Brown PO. Development of the Human Infant Intestinal Microbiota. PLoS Biol 2007; 5:e177.

Pan Q, Mathison J, Fearns C, Kravchenko VV, Da Silva Correia J, Hoffman HM, Kobayashi KS, Bertin J, Grant EP, Coyle AJ, Sutterwala FS, Ogura Y, Flavell RA, Ulevitch RJ. MDP-induced interleukin-1beta processing requires Nod2 and CIAS1/NALP3. J Leukoc Biol 2007; 82: 177-183.

Papin S, Duquesnoy P, Cazeneuve C, Pantel J, Coppey-Moisan M, Dargemont C, Amselem S. Alternative splicing at the MEFV locus involved in familial Mediterranean fever regulates translocation of the marenostrin/pyrin protein to the nucleus. Hum Mol Genet 2000; 9: 3001-3009.

Papin S, Cuenin S, Agostini L, Martinon F, Werner S, Beer HD, Grütter C, Grütter M, Tschopp J. The SPRY domain of Pyrin, mutated in familial Mediterranean fever patients,interacts with inflammasome components and inhibits proIL-1beta processing. Cell Death Differ 2007; 14:1457-1466.

Parham C, Chirica M, Timans J, Vaisberg E, Travis M, Cheung J, Pflanz S, Zhang R, Singh KP, Vega F, To W, Wagner J, O'Farrell AM, McClanahan T, Zurawski S, Hannum C, Gorman D, Rennick DM, Kastelein RA, de Waal Malefyt R, Moore KW. A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rbeta1 and a novel cytokine receptor subunit, IL-23R. J Immunol 2002; 168: 5699-5708.

Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R, Burke DA, Todhunter CE, Ahmad T, Onnie CM, McArdle W, Strachan D, Bethel G, Bryan C, Lewis CM, Deloukas P, Forbes A, Sanderson J, Jewell DP, Satsangi J, Mansfield JC; Wellcome Trust Case Control Consortium, Cardon L, Mathew CG. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 2007; 39: 830-832.

Parry SD, Stansfield R, Jelley D, Gregory W, Phillips E, Barton JR, Welfare MR. Does bacterial gastroenteritis predispose people to functional gastrointestinal disorders? A prospective, community-based, case-control study. Am J Gastroenterol 2003; 98: 1970-1975.

Parry S, Forgacs I. Intestinal infection and irritable bowel syndrome. Eur J Gastroenterol Hepatol 2005; 17: 5-9.

Pasare C, Medzhitov R. Toll-like receptors: linking innate and adaptive immunity. Adv Exp Med Biol 2005; 560: 11-18.

Pasqualetti M, Rijli FM. Homeobox gene mutations and brain-stem developmental disorders: learning from knockout mice. Curr Opin Neurol 2001;14: 177-184.

Pata C, Erdal E, Yazc K, Camdeviren H, Ozkaya M, Ulu O. Association of the -1438 G/A and 102 T/C polymorphism of the 5-Ht2A receptor gene with irritable bowel syndrome 5-Ht2A gene polymorphism in irritable bowel syndrome. J Clin Gastroenterol 2004; 38: 561-566.

Pauleau AL, Murray PJ. Role of nod2 in the response of macrophages to toll-like receptor agonists. Mol Cell Biol 2003; 23: 7531-7539.

Pelegrin P, Surprenant A. Pannexin-1 mediates large pore formation and interleukin-1beta release by the ATP-gated P2X7 receptor. EMBO J. 2006; 25: 5071-5082.

Pelegrin P, Surprenant A. Pannexin-1 couples to maitotoxin- and nigericin-induced interleukin-1beta release through a dye uptake-independent pathway. J Biol Chem. 2007; 282: 2386-2394.

Peltekova VD, Wintle RF, Rubin LA, Amos CI, Huang Q, Gu X, Newman B, Van Oene M, Cescon D, Greenberg G, Griffiths AM, St George-Hyslop PH, Siminovitch KA. Functional variants of OCTN cation transporter genes are associated with Crohn disease. Nat Genet 2004; 36: 471-475.

Perregaux D, Gabel CA. Interleukin-1 beta maturation and release in response to ATP and nigericin. Evidence that potassium depletion mediated by these agents is a necessary and common feature of their activity. J Biol Chem 1994; 269: 15195-15203.

Pétrilli V, Dostert C, Muruve DA, Tschopp J. The inflammasome: a danger sensing complex triggering innate immunity. Curr Opin Immunol 2007a; 19: 615-622.

Pétrilli V, Papin S, Dostert C, Mayor A, Martinon F, Tschopp J. Activation of the NALP3 inflammasome is triggered by low intracellular potassium concentration. Cell Death Differ 2007b; 14: 1583-1589.

Phillips C. Online resources for SNP analysis: a review and route map. Mol Biotechnol 2007; 35: 65-97.

Philpott DJ, Yamaoka S, Israël A, Sansonetti PJ. Invasive Shigella flexneri activates NF-kappa B through a lipopolysaccharide-dependent innate intracellular response and leads to IL-8 expression in epithelial cells. J Immunol 2000; 165: 903-914.

Pizarro-Cerdá J, Cossart P. Bacterial adhesion and entry into host cells. Cell 2006; 124:715-727.

Podolsky DK. Inflammatory bowel disease. N Engl J Med 2002; 347: 417-429.

Powell R. On certain painful afflictions of the intestinal canal. Med Trans R Coll Physicians 1818; 6: 106-117.

Prantera C, Scribano ML, Falasco G, Andreoli A, Luzi C. Ineffectiveness of probiotics in preventing recurrence after curative resection for Crohn's disease: a randomised controlled trial with Lactobacillus GG. Gut 2002; 51: 405– 409.

Pras E, Aksentijevich I, Gruberg L, Balow JE Jr, Prosen L, Dean M, Steinberg AD, Pras M, Kastner DL. Mapping of the gene causing familial Mediterranean fever to the short arm of chromosome 16. N Engl J Med 1992; 326: 1509-1513.

Prieur AM, Griscelli C. Chronic meningo-cutaneo-articular syndrome in children. Rev Rhum Mal Osteoartic. 1980; 47: 645-649.

Prieur AM, Griscelli C. Arthropathy with rash, chronic meningitis, eye lesions, and mental retardation. J Pediatr 1981; 99: 79-83.

Qi L, van Dam RM, Meigs JB, Manson JE, Hunter D, Hu FB. Genetic variation in IL6 gene and type 2 diabetes: tagging-SNP haplotype analysis in large-scale case-control study and meta-analysis. Hum Mol Genet 2006; 15:1914-1920.

Quigley EM. Irritable bowel syndrome and inflammatory bowel disease: interrelated diseases? Chin J Dig Dis 2005; 6:122-132.

Radtke AL, O'Riordan MX. Intracellular innate resistance to bacterial pathogens. Cell Microbiol. 2006; 8: 1720-1729.

Raelson JV, Little RD, Ruether A, Fournier H, Paquin B, Van Eerdewegh P, Bradley WE, Croteau P, Nguyen-Huu Q, Segal J, Debrus S, Allard R,

Rosenstiel P, Franke A, Jacobs G, Nikolaus S, Vidal JM, Szego P, Laplante N, Clark HF, Paulussen RJ, Hooper JW, Keith TP, Belouchi A, Schreiber S. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. Proc Natl Acad Sci U S A 2007; 104: 14747-14752.

Rahimi R, Nikfar S, Rahimi F, Elahi B, Derakhshani S, Vafaie M, Abdollahi M. A meta-analysis on the efficacy of probiotics for maintenance of remission and prevention of clinical and endoscopic relapse in Crohn's disease. Dig Dis Sci 2008; 53: 2524-2531.

Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S, Medzhitov R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. Cell 2004; 118: 229-241.

Ramamoorthy S, Bauman AL, Moore KR, Han H, Yang-Feng T, Chang AS, Ganapathy V, Blakely RD. Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. Proc Natl Acad Sci U S A 1993; 90: 2542-2546.

Raskin DM, Seshadri R, Pukatzki SU, Mekalanos JJ. Bacterial genomics and pathogen evolution. Cell 2006; 124:7 03-714.

Reinhard C, Rioux JD. Role of the IBD5 susceptibility locus in the inflammatory bowel diseases. Inflamm Bowel Dis 2006; 12: 227-238.

Richards N, Schaner P, Diaz A, Stuckey J, Shelden E, Wadhwa A, Gumucio DL. Interaction between pyrin and the apoptotic speck protein (ASC) modulates ASC-induced apoptosis. J Biol Chem 2001; 276: 39320-39329.

Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, Kocher K, Miller K, Guschwan S, Kulbokas EJ, O'Leary S, Winchester E, Dewar K, Green T, Stone V, Chow C, Cohen A, Langelier D, Lapointe G, Gaudet D, Faith J, Branco N, Bull SB, McLeod RS, Griffiths AM, Bitton A, Greenberg GR, Lander ES, Siminovitch KA, Hudson TJ. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 2001; 29: 223-228.

Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 2007; 39: 596-604.

Rodríguez LA, Ruigómez. Increased risk of irritable bowel syndrome after bacterial gastroenteritis: cohort study. BMJ. 1999; 318: 565-566.

Rolfe VE, Fortun PJ, Hawkey CJ, Bath-Hextall F. Probiotics for maintenance of remission in Crohn's disease. Cochrane Database Syst Rev 2006;(4): CD004826.

Rahimi R, Nikfar S, Rahimi F, Elahi B, Derakhshani S, Vafaie M, Abdollahi M. A meta-analysis on the efficacy of probiotics for maintenance of remission and prevention of clinical and endoscopic relapse in Crohn's disease. Dig Dis Sci 2008; 53: 2524-2531.

Russell RK, Satsangi J. IBD: a family affair. Best Pract Res Clin Gastroenterol 2004; 18:525-539.

Russo MW, Gaynes BN, Drossman DA. A national survey of practice patterns of gastroenterologists with comparison to the past two decades. J Clin Gastroenterol 1999; 29: 339-343.

Ryan JG, Kastner DL. Fevers, genes, and innate immunity. Curr Top Microbiol Immunol 2008a; 321: 169-184.

Ryan JG, Goldbach-Mansky R. The spectrum of autoinflammatory diseases: recent bench to bedside observations. Curr Opin Rheumatol 2008b; 20: 66-75.

Saito YA, Locke GR 3rd, Zimmerman JM, Holtmann G, Slusser JP, de Andrade M, Petersen GM, Talley NJ. A genetic association study of 5-HTT LPR and GNbeta3 C825T polymorphisms with irritable bowel syndrome. Neurogastroenterol Motil 2007; 19: 465-470.

Saito YA, Talley NJ. Genetics of Irritable Bowel Syndrome. Am J Gastroenterol 2008a 103:2100-2104.

Saito YA, Zimmerman JM, Harmsen WS, De Andrade M, Locke GR 3rd, Petersen GM, Talley NJ. Irritable bowel syndrome aggregates strongly in families: a family-based case-control study. Neurogastroenterol Motil 2008b; 2: 790-797.

Samuels J, Aksentijevich I, Torosyan Y, Centola M, Deng Z, Sood R, Kastner DL. Familial Mediterranean fever at the millennium. Clinical spectrum, ancient mutations, and a survey of 100 American referrals to the National Institutes of Health. Medicine (Baltimore). 1998; 77: 268-297.

Sandler RS, Everhart JE, Donowitz M, Adams E, Cronin K, Goodman C, Gemmen E, Shah S, Avdic A, Rubin R. The burden of selected digestive diseases in the United States. Gastroenterology 2002; 122 (5):1500-1511.

Sari S, Egritas O, Dalgic B. The familial Mediterranean fever (MEFV) gene may be a modifier factor of inflammatory bowel disease in infancy. Eur J Pediatr. 2008; 167: 391-393.

Sarkar A, Duncan M, Hart J, Hertlein E, Guttridge DC, Wewers MD. ASC directs NF-kappaB activation by regulating receptor interacting protein-2 (RIP2) caspase-1 interactions. J Immunol 2006; 176: 4979-4986.

Sartor RB. Microbial influences in inflammatory bowel diseases. Gastroenterology 2008; 134: 577-594.

Sasaki M, Sitaraman SV, Babbin BA, Gerner-Smidt P, Ribot EM, Garrett N, Alpern JA, Akyildiz A, Theiss AL, Nusrat A, Klapproth JM. Invasive Escherichia coli are a feature of Crohn's disease. Lab Invest 2007; 87:1042-1054.

Satsangi J, Welsh KI, Bunce M, Julier C, Farrant JM, Bell JI, Jewell DP. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet 1996; 347: 1212-1217.

Sauer H, Hescheler J, Wartenberg M. Mechanical strain-induced Ca(2+) waves are propagated via ATP release and purinergic receptor activation. Am J Physiol Cell Physiol 2000; 279: C295-307.

Savage DC. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 1977; 31: 107-33.

Schoenborn JR, Wilson CB. Regulation of interferon-gamma during innate and adaptive immune responses. Adv Immunol 2007; 96: 41-101.

Schreiber S, Heinig T, Thiele HG, Raedler A. Immunoregulatory role of interleukin 10 in patients with inflammatory bowel disease. Gastroenterology 1995; 108: 1434-4144.

Schmid D, Dengjel J, Schoor O, Stevanovic S, Münz C. Autophagy in innate and adaptive immunity against intracellular pathogens. J Mol Med 2006; 84: 194-202.

Schmidt C, Giese T, Ludwig B, Mueller-Molaian I, Marth T, Zeuzem S, Meuer SC, Stallmach A. Expression of interleukin-12-related cytokine transcripts in inflammatory bowel disease: elevated interleukin-23p19 and interleukin-27p28 in Crohn's disease but not in ulcerative colitis. Inflamm Bowel Dis 2005; 11: 16-23.

Sellon RK, Tonkonogy S, Schultz M, Dieleman LA, Grenther W, Balish E, Rennick DM, Sartor RB. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. Infect Immun 1998; 66: 5224-5231.

Seshadri S, Duncan MD, Hart JM, Gavrilin MA, Wewers MD. Pyrin levels in human monocytes and monocyte-derived macrophages regulate IL-1beta processing and release. J Immunol 2007; 179: 1274-1281.

Shaheen NJ, Hansen RA, Morgan DR, Gangarosa LM, Ringel Y, Thiny MT, Russo MW, Sandler RS. The burden of gastrointestinal and liver diseases, 2006. Am J Gastroenterol 2006; 101: 2128-2138.

Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res 1999; 9: 677-679.

Shi Y, Evans JE, Rock KL. Molecular identification of a danger signal that alerts the immune system to dying cells. Nature 2003; 425: 516-521.

Shih YC, Barghout VE, Sandler RS, Jhingran P, Sasane M, Cook S, Gibbons DC, Halpern M. Resource utilization associated with irritable bowel syndrome in the United States 1987-1997. Dig Dis Sci 2002; 47: 1705-1715.

Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, Caprilli R, Colombel JF, Gasche C, Geboes K, Jewell DP, Karban A, Loftus Jr EV, Peña AS, Riddell RH, Sachar DB, Schreiber S, Steinhart AH, Targan SR, Vermeire S, Warren BF. 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): 5-36.

Silverberg MS. OCTNs: will the real IBD5 gene please stand up? World J Gastroenterol 2006; 12: 3678-3681.

Silverberg MS, Duerr RH, Brant SR, Bromfield G, Datta LW, Jani N, Kane SV, Rotter JI, Philip Schumm L, Hillary Steinhart A, Taylor KD, Yang H, Cho JH, Rioux JD, Daly MJ; NIDDK IBD Genetics Consortium. Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. Eur J Hum Genet 2007; 15: 328-335.

Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, Achkar JP, Goyette P, Scott R, Xu W, Barmada MM, Klei L, Daly MJ, Abraham C,

Bayless TM, Bossa F, Griffiths AM, Ippoliti AF, Lahaie RG, Latiano A, Paré P, Proctor DD, Regueiro MD, Steinhart AH, Targan SR, Schumm LP, Kistner EO, Lee AT, Gregersen PK, Rotter JI, Brant SR, Taylor KD, Roeder K, Duerr RH. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet 2009 (*in press*).

Simrén M, Axelsson J, Gillberg R, Abrahamsson H, Svedlund J, Björnsson ES. Quality of life in inflammatory bowel disease in remission: the impact of IBS-like symptoms and associated psychological factors. Am J Gastroenterol 2002; 97: 389-396.

Singh SB, Davis AS, Taylor GA, Deretic V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. Science 2006; 313: 1438-1441.

Sohar E, Gafni J, Pras M, Heller H. Familial Mediterranean fever. A survey of 470 cases and review of the literature. Am J Med 1967; 43: 227-253.

Solle M, Labasi J, Perregaux DG, Stam E, Petrushova N, Koller BH, Griffiths RJ, Gabel CA. Altered cytokine production in mice lacking P2X(7) receptors. J Biol Chem 2001; 276: 125-132.

Sonnenburg JL, Angenent LT, Gordon JI. Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? Nat Immunol 2004; 5: 569-573.

Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993; 52: 506-516.

Spiller RC, Jenkins D, Thornley JP, Hebden JM, Wright T, Skinner M, Neal KR. Increased rectal mucosal enteroendocrine cells, T lymphocytes, and increased gut permeability following acute Campylobacter enteritis and in post-dysenteric irritable bowel syndrome. Gut 2000; 47: 804-811.

Spiller RC. Postinfectious irritable bowel syndrome. Gastroenterology 2003a; 124:1662-1671.

Spiller RC. Estimating the importance of infection in IBS. Am J Gastroenterol 2003b; 98: 238-241.

Spiller R, Campbell E. Post-infectious irritable bowel syndrome. Curr Opin Gastroenterol 2006; 22: 13-17.

Spiller RC. Role of infection in irritable bowel syndrome. J Gastroenterol 2007a; 42 (Suppl 17): 41-47.

Spiller RC. Is IBS caused by infectious diarrhea? Nat Clin Pract Gastroenterol Hepatol 2007b; 4: 642-643.

Srinivasula SM, Poyet JL, Razmara M, Datta P, Zhang Z, Alnemri ES. The PYRIN-CARD protein ASC is an activating adaptor for caspase-1. J Biol Chem 2002; 277: 21119-21122.

Stappenbeck TS, Hooper LV, Gordon JI. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. Proc Natl Acad Sci USA 2002; 99: 15451-15455.

Stark PL, Lee A. The microbial ecology of the large bowel of breast-fed and formula-fed infant during the first year of life. J Med Microbiol. 1982 May;15(2):189-203.

Stehlik C, Fiorentino L, Dorfleutner A, Bruey JM, Ariza EM, Sagara J, Reed JC. The PAAD/PYRIN-family protein ASC is a dual regulator of a conserved step in nuclear factor kappaB activation pathways. J Exp Med. 2002; 196: 1605-1615.

Stemmler MP, Hecht A, Kemler R. E-cadherin intron 2 contains cis-regulatory elements essential for gene expression. Development 2005; 132:965-976.

Stojanov S, Kastner DL. Familial autoinflammatory diseases: genetics, pathogenesis and treatment. Curr Opin Rheumatol 2005; 17: 586-599.

Stokkers PC, Reitsma PH, Tytgat GN, van Deventer SJ. HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. Gut 1999; 45: 395-401.

Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA, Rosenstiel P, Albrecht M, Croucher PJ, Seegert D, Nikolaus S, Hampe J, Lengauer T, Pierrou S, Foelsch UR, Mathew CG, Lagerstrom-Fermer M, Schreiber S. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nat Genet 2004; 36: 476-480.

Sugimoto K, Ogawa A, Mizoguchi E, Shimomura Y, Andoh A, Bhan AK, Blumberg RS, Xavier RJ, Mizoguchi A. IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. J Clin Invest 2008; 118: 534-544.

Suh CI, Stull ND, Li XJ, Tian W, Price MO, Grinstein S, Yaffe MB, Atkinson S, Dinauer MC. The phosphoinositide-binding protein p40phox activates the NADPH oxidase during FcgammaIIA receptor-induced phagocytosis. J Exp Med 2006; 203: 1915-1925.

Sutterwala FS, Ogura Y, Szczepanik M, Lara-Tejero M, Lichtenberger GS, Grant EP, Bertin J, Coyle AJ, Galán JE, Askenase PW, Flavell RA. Critical role for NALP3/CIAS1/Cryopyrin in innate and adaptive immunity through its regulation of caspase-1. Immunity 2006; 24: 317-327.

Suzuki N, Ishizaki J, Yokota Y, Higashino K, Ono T, Ikeda M, Fujii N, Kawamoto K, Hanasaki K. Structures, enzymatic properties, and expression of novel human and mouse secretory phospholipase A(2)s. J Biol Chem 2000; 275: 5785-5793.

Swidsinki A, Khilkin M, Ortner M. Alteration of bacterial concentrations in colonic biopsies from patients with irritable bowel syndrome. Gastroenterology 1999; 116: A1.

Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 2002; 3:391-397.

Talley NJ, Phillips SF, Melton J, Wiltgen C, Zinsmeister AR. A patient questionnaire to identify bowel disease. Ann Intern Med 1989; 111: 671–674.

Talley NJ. Irritable bowel syndrome. Intern Med J 2006; 36: 724-728.

Targan SR, Hanauer SB, van Deventer SJ, Mayer L, Present DH, Braakman T, DeWoody KL, Schaible TF, Rutgeerts PJ. A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease cA2 Study Group. N Engl J Med 1997; 337: 1029–1035.

Taylor GA. IRG proteins: key mediators of interferon-regulated host resistance to intracellular pathogens. Cell Microbiol 2007; 9: 1099-1107.

Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, Chiaromonte F. ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Res 2006; 16: 1596-1604.

Thabane M, Kottachchi DT, Marshall JK. Systematic review and meta-analysis: The incidence and prognosis of post-infectious irritable bowel syndrome. Aliment Pharmacol Ther 2007; 26:535-544.

The French FMF Consortium. A candidate gene for familial Mediterranean fever. Nat Genet 1997; 17: 25-31.

The International FMF Consortium. Ancient missense mutations in a new member of the RoRet gene family are likely to cause familial Mediterranean fever. Cell 1997; 90: 797-807.

Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Genetics versus environment in inflammatory bowel disease: results of a British twin study. BMJ 1996; 312: 95-96.

Thompson WG, Creed F, Drossman DA, Heaton KW, Mazzacca G. Functional bowel disorders and functional abdominal pain. Gastroenterol Int 1992; 5:75–91.

Thompson WG, Longstreth GF, Drossman DA, Heaton KW, Irvine EJ, Müller-Lissner SA. Functional bowel disorders and functional abdominal pain. Gut 1999; 45 Suppl 2: II43-47.

Thornberry NA, Bull HG, Taub D, Wilson KE, Giménez-Gallego G, Rosegay A, Soderman DD, Patchett AA. Mechanism-based inactivation of alanine racemase by 3-halovinylglycines. J Biol Chem 1991; 266: 21657-21665.

Thornley JP, Jenkins D, Neal K, Wright T, Brough J, Spiller RC. Relationship of Campylobacter toxigenicity in vitro to the development of postinfectious irritable bowel syndrome. J Infect Dis 2001; 184:606-609.

Thornton T, McPeek MS. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. Am J Hum Genet 2007; 81: 321-337.

Tidow N, Chen X, Müller C, Kawano S, Gombart AF, Fischel-Ghodsian N, Koeffler HP. Hematopoietic-specific expression of MEFV, the gene mutated in familial Mediterranean fever, and subcellular localization of its corresponding protein, pyrin. Blood 2000; 95: 1451-1455.

Ting JP, Kastner DL, Hoffman HM. CATERPILLERs, pyrin and hereditary immunological disorders. Nat Rev Immunol 2006; 6: 183–195.

Tremelling M, Waller S, Bredin F, Greenfield S, Parkes M. Genetic variants in TNF-alpha but not DLG5 are associated with inflammatory bowel disease in a large United Kingdom cohort. Inflamm Bowel Dis 2006; 12: 178-184.

Tschopp J, Irmler M, Thome M. Inhibition of fas death signals by FLIPs. Curr Opin Immunol 1998; 10: 552-558.

Tunca M, Ozdogan H. Molecular and genetic characteristics of hereditary autoinflammatory diseases. Curr Drug Targets Inflamm Allergy 2005; 4: 77-80.

Turner JR. Molecular basis of epithelial barrier regulation: from basic mechanisms to clinical application. Am J Pathol 2006; 169: 1901-1909.

Tysk C, Lindberg E, Järnerot G, Flodérus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. Gut 1988; 29: 990-996.

Uhlig HH, McKenzie BS, Hue S, Thompson C, Joyce-Shaikh B, Stepankova R, Robinson N, Buonocore S, Tlaskalova-Hogenova H, Cua DJ, Powrie F. Differential activity of IL-12 and IL-23 in mucosal and systemic innate immune pathology. Immunity 2006; 25: 309-318.

van Beelen AJ, Zelinkova Z, Taanman-Kueter EW, Muller FJ, Hommes DW, Zaat SA, Kapsenberg ML, de Jong EC. Stimulation of the intracellular bacterial sensor NOD2 programs dendritic cells to promote interleukin-17 production in human memory T cells. Immunity 2007; 27: 660-669.

van Bodegraven AA, Curley CR, Hunt KA, Monsuur AJ, Linskens RK, Onnie CM, Crusius JB, Annese V, Latiano A, Silverberg MS, Bitton A, Fisher SA, Steinhart AH, Forbes A, Sanderson J, Prescott NJ, Strachan DP, Playford RJ, Mathew CG, Wijmenga C, Daly MJ, Rioux JD, van Heel DA. Genetic variation in myosin IXB is associated with ulcerative colitis. Gastroenterology. 2006; 131: 1768-1774.

van den Boom D, Ehrich M. Discovery and identification of sequence polymorphisms and mutations with MALDI-TOF MS. Methods Mol Biol 2007; 366: 287-306.

van der Veek PP, van den Berg M, de Kroon YE, Verspaget HW, Masclee AA. Role of tumor necrosis factor-alpha and interleukin-10 gene polymorphisms in irritable bowel syndrome. Am J Gastroenterol 2005; 100: 2510-2516.

van Duist MM, Albrecht M, Podswiadek M, Giachino D, Lengauer T, Punzi L, De Marchi M. A new CARD15 mutation in Blau syndrome. Eur J Hum Genet 2005; 13: 742-747.

van Dullemen HM, van Deventer SJ, Hommes DW, Bijl HA, Jansen J, Tytgat GN, Woody J. Treatment of Crohn's disease with anti-tumor necrosis factor chimeric monoclonal antibody (cA2). Gastroenterology 1995; 109: 129–135.

Van Gossum A, Dewit O, Louis E, de Hertogh G, Baert F, Fontaine F, DeVos M, Enslen M, Paintin M, Franchimont. Multicenter randomized-controlled clinical trial of probiotics (Lactobacillus johnsonii, LA1) on early endoscopic recurrence of Crohn's disease after lleo caecal resection. Inflamm Bowel Dis 2007; 13:135-142.

van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM; Genome Scan Meta-Analysis Group of the IBD International Genetics Consortium. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. Hum Mol Genet. 2004; 13: 763-770.

van Heel DA, Ghosh S, Butler M, Hunt KA, Lundberg AM, Ahmad T, McGovern DP, Onnie C, Negoro K, Goldthorpe S, Foxwell BM, Mathew CG, Forbes A, Jewell DP, Playford RJ. Muramyl dipeptide and toll-like

receptor sensitivity in NOD2-associated Crohn's disease. Lancet 2005; 365: 1794-1796.

van Heel DA, Hunt KA, King K, Ghosh S, Gabe SM, Mathew CG, Forbes A, Playford RJ. Detection of muramyl dipeptide-sensing pathway defects in patients with Crohn's disease. Inflamm Bowel Dis 2006; 12: 598-605.

Van Kerkhoven LA, Laheij RJ, Jansen JB. Meta-analysis: a functional polymorphism in the gene encoding for activity of the serotonin transporter protein is not associated with the irritable bowel syndrome. Aliment Pharmacol Ther 2007; 26: 979-986.

Vasilakos JP, Ghayur T, Carroll RT, Giegel DA, Saunders JM, Quintal L, Keane KM, Shivers BD. IL-1 beta converting enzyme (ICE) is not required for apoptosis induced by lymphokine deprivation in an IL-2-dependent T cell line. J Immunol 1995; 155: 3433-3442.

Verdú EF, Bercik P, Verma-Gandhu M, Huang XX, Blennerhassett P, Jackson W, Mao Y, Wang L, Rochat F, Collins SM. Specific probiotic therapy attenuates antibiotic induced visceral hypersensitivity in mice. Gut 2006; 55: 182-190.

Villani AC, Lemire M, Fortin G, Louis E, Silverberg MS, Collette C, Baba N, Libioulle C, Belaiche J, Bitton A, Gaudet D, Cohen A, Langelier D, Fortin PR, Wither JE, Sarfati M, Rutgeerts P, Rioux JD, Vermeire S, Hudson TJ, Franchimont D. Common variants in the NLRP3 region contribute to Crohn disease susceptibility. Nat Genet 2009; 41: 71-76.

Wagner H. Bacterial CpG DNA activates immune cells to signal infectious danger. Adv Immunol 1999; 73: 329-368.

Walev I, Reske K, Palmer M, Valeva A, Bhakdi S. Potassium-inhibited processing of IL-1 beta in human monocytes. EMBO J 1995; 14: 1607-1614.

Wang L, Manji GA, Grenier JM, Al-Garawi A, Merriam S, Lora JM, Geddes BJ, Briskin M, DiStefano PS, Bertin J. PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates activation of NF-kappa B and caspase-1-dependent cytokine processing. J Biol Chem 2002; 277: 29874-2980.

Wang LH, Fang XC, Pan GZ. Bacillary dysentery as a causative factor of irritable bowel syndrome and its pathogenesis. Gut. 2004; 53: 1096-1101.

Watanabe T, Kitani A, Murray PJ, Wakatsuki Y, Fuss IJ, Strober W. Nucleotide binding oligomerization domain 2 deficiency leads to dysregulated TLR2 signaling and induction of antigen-specific colitis. Immunity 2006; 25: 473-485.

Watanabe H, Gaide O, Pétrilli V, Martinon F, Contassot E, Roques S, Kummer JA, Tschopp J, French LE. Activation of the IL-1beta-processing inflammasome is involved in contact hypersensitivity. J Invest Dermatol 2007; 127: 1956-1963.

Watford WT, Hissong BD, Bream JH, Kanno Y, Muul L, O'Shea JJ. Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4. Immunol Rev 2004; 202: 139-156.

Wehkamp J, Salzman NH, Porter E, Nuding S, Weichenthal M, Petras RE, Shen B, Schaeffeler E, Schwab M, Linzmeier R, Feathers RW, Chu H, Lima H Jr, Fellermann K, Ganz T, Stange EF, Bevins CL. Reduced Paneth cell

alpha-defensins in ileal Crohn's disease. Proc Natl Acad Sci USA 2005; 102: 18129-18134.

Wehkamp J, Stange EF. Paneth cells and the innate immune response. Curr Opin Gastroenterol 2006; 22: 644-650.

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007a; 447: 661-678.

Wellcome Trust Case Control Consortium; Australo-Anglo-American Spondylitis Consortium (TASC), Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007b; 39: 1329-1337.

Werts C, Girardin SE, Philpott DJ. TIR, CARD and PYRIN: three domains for an antimicrobial triad. Cell Death Differ 2006; 13: 798-815.

Weston AP, Biddle WL, Bhatia PS, Miner PB Jr. Terminal ileal mucosal mast cells in irritable bowel syndrome. Dig Dis Sci 1993; 38: 1590-1595.

Whitcombe D, Brownie J, Gillard HL, McKechnie D, Theaker J, Newton CR, Little S. A homogeneous fluorescence assay for PCR amplicons: its application to real-time, single-tube genotyping. Clin Chem 1998; 44: 918-923.

Whitehead WE, Bosmajian L, Zonderman AB, Costa PT Jr, Schuster MM. Symptoms of psychologic distress associated with irritable bowel syndrome. Comparison of community and medical clinic samples. Gastroenterology 1988; 95: 709-714.

Whorwell PJ. Serotonin receptor modulation in irritable bowel syndrome: one step forwards and one step backwards. Gut 2001; 48: 596-597.

Wiekowski MT, Leach MW, Evans EW, Sullivan L, Chen SC, Vassileva G, Bazan JF, Gorman DM, Kastelein RA, Narula S, Lira SA. Ubiquitous transgenic expression of the IL-23 subunit p19 induces multiorgan inflammation, runting, infertility, and premature death. J Immunol 2001; 166: 7563-7570.

Wilks S. Morbid appearances in the intestines of Miss Banks. Medical Times & Gazette 1859; 2: 264-265.

Williams CN, Kocher K, Lander ES, Daly MJ, Rioux JD. Using a genome-wide scan and meta-analysis to identify a novel IBD locus and confirm previously identified IBD loci. Inflamm Bowel Dis 2002; 8: 375-381.

Woo JS, Imm JH, Min CK, Kim KJ, Cha SS, Oh BH. Structural and functional insights into the B30.2/SPRY domain. EMBO J 2006; 25: 1353-1363.

Wu S, Rhee KJ, Zhang M, Franco A, Sears CL. Bacteroides fragilis toxin stimulates intestinal epithelial cell shedding and gamma-secretase-dependent E-cadherin cleavage. J Cell Sci 2007; 120:1944-1952.

Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. Nature 2007; 448: 427-434.

Xavier RJ, Huett A, Rioux JD. Autophagy as an important process in gut homeostasis and Crohn's disease pathogenesis. Gut 2008; 57: 717-720.

Xu H, Shete S. Mixed-effects logistic approach for association following linkage scan for complex disorders. Ann Hum Genet 2007; 71: 230-237.

Yamazaki K, Takazoe M, Tanaka T, Kazumori T, Nakamura Y. Absence of mutation in the NOD2/CARD15 gene among 483 Japanese patients with Crohn's disease. J Hum Genet 2002; 47: 469-472.

Yamazaki K, McGovern D, Ragoussis J, Paolucci M, Butler H, Jewell D, Cardon L, Takazoe M, Tanaka T, Ichimori T, Saito S, Sekine A, Iida A, Takahashi A, Tsunoda T, Lathrop M, Nakamura Y. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. Hum Mol Genet 2005; 14: 3499-3506.

Yamazaki K, Onouchi Y, Takazoe M, Kubo M, Nakamura Y, Hata A. Association analysis of genetic variants in IL23R, ATG16L1 and 5p13.1 loci with Crohn's disease in Japanese patients. J Hum Genet 2007; 52: 575-583.

Yang D, Chen Q, Yang H, Tracey KJ, Bustin M, Oppenheim JJ. High mobility group box-1 protein induces the migration and activation of human dendritic cells and acts as an alarmin. J Leukoc Biol 2007; 81: 59-66.

Yen D, Cheung J, Scheerens H, Poulet F, McClanahan T, McKenzie B, Kleinschek MA, Owyang A, Mattson J, Blumenschein W, Murphy E, Sathe M, Cua DJ, Kastelein RA, Rennick D. IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. J Clin Invest 2006; 116: 1310-1316.

Yu HB, Finlay BB. The caspase-1 inflammasome: a pilot of innate immune responses. Cell Host Microbe 2008; 4: 198-208.

Yu JW, Wu J, Zhang Z, Datta P, Ibrahimi I, Taniguchi S, Sagara J, Fernandes-Alnemri T, Alnemri ES. Cryopyrin and pyrin activate caspase-1, but not

NF-kappaB, via ASC oligomerization. Cell Death Differ 2006; 13: 236-249.

Zaykin DV, Shibata K. Genetic flip-flop without an accompanying change in linkage disequilibrium. Am J Hum Genet 2008; 82: 794-796.

Zeggini E, Rayner W, Morris AP, Hattersley AT, Walker M, Hitman GA, Deloukas P, Cardon LR, McCarthy MI. An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. Nat Genet 2005; 37: 1320-1322.

Zheng JJ, Zhu XS, Huangfu Z, Gao ZX, Guo ZR, Wang Z. Crohn's disease in mainland China: a systematic analysis of 50 years of research. Chin J Dig Dis 2005; 6: 175-181.

Zoetendal EG, Akkermans AD, De Vos WM. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. Appl Environ Microbiol 1998; 64: 3854-3859.

# 9 ELECTRONIC SOURCES

- HapMap project database, http://www.hapmap.org/
- UCSC genome browser (Golden Path), http://genome.ucsc.edu/
- NCBI, http://www.ncbi.nlm.nih.gov/
- Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim
- dbSNP database, http://www.ncbi.nlm.nih.gov/projects/SNP/
- Ensembl, http://www.ensembl.org/index.html
- HUGO gene nomenclature Committee, http://www.genenames.org/
- INFEVERS database, http://fmf.igh.cnrs.fr/ISSAID/infevers/
- Primer 3 softwer, http://frodo.wi.mit.edu/
- ESPERR software, http://www.bx.psu.edu/projects/esperr
- Tagger server, http://www.broad.mit.edu/mpg/tagger/server.html
- Single Nucleotide Polymorphism Spectral Decomposition (SNPSpD) software, http://gump.qimr.edu.au/general/daleN/SNPSpD/
- R software: http://www.r-project.org/
- 1000 genome project, http://www.1000genomes.org/

# 10 APPENDIX

- Print copy of the letter published in *Nature Genetics*; reference:
    - Villani AC, Lemire M, Fortin G, Louis E, Silverberg MS, Collette C, Baba N, Libioulle C, Belaiche J, Bitton A, Gaudet D, Cohen A, Langelier D, Fortin PR, Wither JE, Sarfati M, Rutgeerts P, Rioux JD, Vermeire S, Hudson TJ, Franchimont D. Common variants in the NLRP3 region contribute to Crohn disease susceptibility. *Nat Genet* 2009; 41: 71-76.

- Copy of the license to publish
- Co-authors' copyright waiver forms (i.e. copies of agreement through e-mail exchange) for unpublished material submitted for publication
- McGill University environmental health and safety certificate
- Ethical approvals

nature genetics

# Common variants in the *NLRP3* region contribute to Crohn's disease susceptibility

Alexandra-Chloé Villani[1,2], Mathieu Lemire[3], Geneviève Fortin[1], Edouard Louis[4], Mark S Silverberg[5], Catherine Collette[1], Nobuyasu Baba[6], Cécile Libioulle[4], Jacques Belaiche[4], Alain Bitton[7], Daniel Gaudet[8], Albert Cohen[9], Diane Langelier[10], Paul R Fortin[11], Joan E Wither[11], Marika Sarfati[6], Paul Rutgeerts[12], John D Rioux[13], Severine Vermeire[12], Thomas J Hudson[3] & Denis Franchimont[1,14]



**We used a candidate gene approach to identify a set of SNPs, located in a predicted regulatory region on chromosome 1q44 downstream of *NLRP3* (previously known as *CIAS1* and *NALP3*) that are associated with Crohn's disease. The associations were consistently replicated in four sample sets from individuals of European descent. In the combined analysis of all samples (710 father-mother-child trios, 239 cases and 107 controls), these SNPs were strongly associated with risk of Crohn's disease ($P_{combined} = 3.49 \times 10^{-9}$, odds ratio = 1.78, confidence interval = 1.47–2.16 for rs10733113), reaching a level consistent with the stringent significance thresholds imposed by whole-genome association studies. In addition, we observed significant associations between SNPs in the associated regions and *NLRP3* expression and IL-1β production. Mutations in *NLRP3* are known to be responsible for three rare autoinflammatory disorders[1,2]. These results suggest that the *NLRP3* region is also implicated in the susceptibility of more common inflammatory diseases such as Crohn's disease.**

Crohn's disease and ulcerative colitis are multigenic and heterogeneous inflammatory bowel diseases of the gastrointestinal tract that seem to result from a dysregulated mucosal immune response to bacterial antigens in the gut lumen of a genetically susceptible host[3]. *NLRP3* is a member of the CATERPILLER[4] family of genes encoding proteins that comprise a nucleotide-binding domain and a leucine-rich repeat domain. Cryopyrin, the protein encoded by *NLRP3*, controls the inflammasome, a crucial molecular platform that regulates activation of caspase-1 and processing of interleukin (IL)-1β—two key mediators of inflammation[2,5,6]. The potential involvement of *NLRP3* in the pathogenesis of more common inflammatory disorders motivated us to conduct an in-depth genetic analysis of the *NLRP3* region.

We first assessed the association between 47 SNPs in the *NLRP3* region and Crohn's disease risk in 296 trios from Leuven University Hospital Gasthuisberg (see Methods, **Supplementary Methods** and **Supplementary Table 1** online). The major alleles of three tagging SNPs were significantly associated with increased risk of Crohn's disease in the Leuven sample set ($P = 0.0107$ for rs4353135; $P = 7.63 \times 10^{-3}$ for rs4266924; $P = 1.68 \times 10^{-3}$ for rs10733113; **Fig. 1** and **Table 1**). These SNPs span a 5.3-kb region and are located 4.7 kb downstream of *NLRP3* and 1.85 kb upstream of the olfactory receptor gene *OR2B11* (**Fig. 1**). No associations were observed with SNPs within the *NLRP3* gene, and none of the associated SNPs were in linkage disequilibrium with tagging SNPs located in *NLRP3* or *OR2B11*. These preliminary associations pointed to the 5.3-kb region near *NLRP3*, bounded by rs4353135 and rs10733113, as a candidate region contributing to Crohn's disease susceptibility.

In the second phase of the study, we examined the association of the above three SNPs with risk of Crohn's disease in additional samples. Overall, we screened one case-control cohort (Liège) and three familial sample sets (Liège, Québec and Toronto; see Methods and **Supplementary Methods**). We replicated our initial significant Crohn's disease associations with the major alleles of tagging SNPs rs4266924 and rs10733113 in all four sample sets ($P < 0.05$; **Fig. 1** and **Table 1**). For the Québec and Toronto cohorts, the observed associations

[1]Division of Gastroenterology, McGill University Health Centre, Montreal General Hospital, 1650 Cedar Avenue, Montréal, Québec H3G 1A4, Canada. [2]McGill University and Génome Québec Innovation Centre, 740 Doctor Penfield Avenue, Montréal, Québec H3A 1A4, Canada. [3]Ontario Institute for Cancer Research, MaRS Centre, South Tower, 101 College Street, Suite 800, Toronto, Ontario M5G 0A3, Canada. [4]Division of Gastroenterology, CHU, Université de Liège, 13 avenue de l'Hôpital Bât. B35, 4000 Liège, Belgium. [5]Mount Sinai Hospital IBD Centre, University of Toronto, 441-600 University Avenue, Toronto, Ontario M5G 1X5, Canada. [6]Immunoregulation Laboratory, CHUM Research Center, Université de Montréal, 1560 rue Sherbrooke Est, Montréal, Québec H2L 4M1, Canada. [7]Division of Gastroenterology, McGill University Health Centre, Royal Victoria Hospital, 687 Pine Avenue West, Montréal, Québec H3A 1A1, Canada. [8]Université de Montréal, Centre de médecine génique communautaire de l'Université de Montréal, Hôpital universitaire de Chicoutimi, 555 boulevard de l'Université, Chicoutimi, Québec G7H 2B1, Canada. [9]Jewish General Hospital, 3755 chemin de la Côte-Sainte-Catherine, Montréal, Québec H3T 1E2, Canada. [10]Centre Hospitalier Universitaire de Sherbrooke–Hôtel-Dieu, 580 rue Bowen Sud, Sherbrooke, Québec J1G 2E8, Canada. [11]Arthritis Centre of Excellence, Toronto Western Hospital Research Institute, 399 Bathurst Street, University of Toronto, Toronto, Ontario M5T 2S8, Canada. [12]Department of Gastroenterology, University Hospital Gasthuisberg, 49 Herestraat, 3000 Leuven, Belgium. [13]Université de Montréal and Montréal Heart Institute, Research Center, 5000 rue Bélanger, Montréal, Québec H1T 1C8, Canada. [14]Department of Gastroenterology, Erasme Hospital, ULB, 808 Lennik road, 1070 Brussels, Belgium. Correspondence should be addressed to A.-C.V. (alexandra-chloe.villani@mail.mcgill.ca) or D.F. (Denis.Franchimont@erasme.ulb.ac.be).

Received 21 July; accepted 2 October; published online 21 December 2008; doi:10.1038/ng.285

remained whether the analysis was done with or without samples from individuals of Jewish ancestry (data not shown). Association with rs4353135 was replicated only in the Toronto sample set ($P < 0.05$; **Fig. 1** and **Table 1**). Combined analysis of all Crohn's disease samples revealed strong associations for rs4353135 ($P_{combined} = 8.36 \times 10^{-3}$, odds ratio = 1.21, confidence interval = 1.05–1.39, T allele frequency = 71% in cases and 65% in controls), rs4266924 ($P_{combined} = 6.01 \times 10^{-7}$, odds ratio = 1.69, confidence interval = 1.37–2.07, A allele frequency = 91% in cases and 84% in controls) and rs10733113 ($P_{combined} = 3.49 \times 10^{-9}$, odds ratio = 1.78, confidence interval = 1.47–2.16, G allele frequency = 90% in cases and 80% in controls).

This regional association was the first step in localizing the most likely causal variant(s). Because the observed association signal was not in linkage disequilibrium with any genotyped variant within *NLRP3*, we resequenced a 9-kb region extending from the *NLRP3* 3′ UTR to the 5.3-kb region described above, inclusively, which also comprised *OR2B11* (**Supplementary Table 2** online). Overall, we selected 16 Crohn's disease samples and 8 controls based on genotypes at markers rs4266924 and rs10733113 to fully define the linkage disequilibrium within the region and identify all polymorphisms in linkage disequilibrium with the associated SNPs. The resequencing effort identified 79 SNPs, 8 of which were previously genotyped. Among these SNPs, 14 were novel compared to dbSNP release 129, and 60 had a minor allele frequency ≥ 0.05 (**Supplementary Table 3** online).

We next conducted comprehensive genotyping to identify polymorphisms with stronger associations (**Supplementary Methods**). We genotyped a total of 24 SNPs in the Leuven exploratory Crohn's disease trios (**Supplementary Table 3**) and analyzed 15 of them after they passed quality control tests (see Methods). None of the SNPs within *OR2B11* (**Supplementary Table 3**) were associated with Crohn's disease. The major alleles of three SNPs, spanning a 1.8-kb region bounded by rs4353135 and rs10733113, were associated with Crohn's disease (**Fig. 1** and **Table 1**). According to the computational method ESPERR[7], two of these SNPs (rs6672995 and rs55646866) are located in a predicted regulatory region. The third variant, ss107635144, was selected by the tagging algorithm[8]. These SNPs were in high linkage disequilibrium ($r^2 > 0.70$) with rs4266924 and rs10733113. None of these three SNPs were tags for the SNPs that were not genotyped in the region.

We subsequently screened these three variants in the other four sample sets, and their associations were consistently replicated (**Fig. 1** and **Table 1**). Combined analysis of all Crohn's disease samples revealed strong associations for rs55646866 ($P_{combined} = 7.2 \times 10^{-7}$, odds ratio = 1.69, confidence interval = 1.38–2.08, C allele frequency = 91% in cases and 84% in controls), rs6672995 ($P_{combined} = 2.91 \times 10^{-6}$, odds ratio = 1.53, confidence interval = 1.28–1.82, G allele frequency = 87% in cases and 79% in controls) and ss107635144 ($P_{combined} = 8.50 \times 10^{-6}$, odds ratio = 1.53, confidence interval = 1.27–1.84, C allele frequency = 88% in cases and 80% in controls).

Conditioning on one of these associated SNPs to evaluate the residual significance of the others did not provide evidence of additive effects of the associated SNPs on the risk. Rather, the association they all showed with Crohn's disease can be explained solely by linkage disequilibrium. To account for multiple testing issues, we note that if all 62 SNPs had been genotyped in the Leuven exploratory sample in the same study phase, then rs55646866 would have reached a significance level ($P = 5.72 \times 10^{-4}$) that, after correction using a permutation procedure, would have still been significant ($P_{corrected} = 0.019$, estimated from 5,000 replicates; see Methods). This result was further strengthened by the consistent replication of rs55646866 in all studied samples.

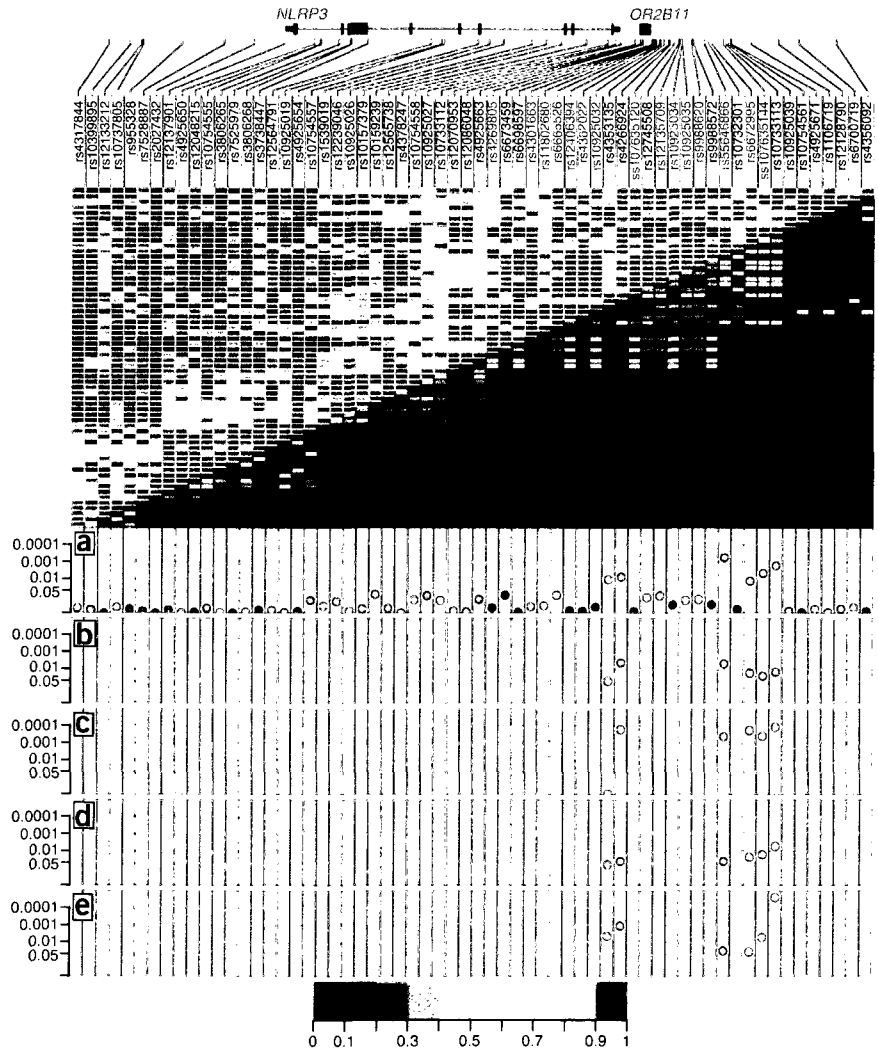



**Figure 1** Association results for the five Crohn's disease sample sets. Top panel shows SNPs, their positions in the genes and the linkage disequilibrium structure between them. SNP names in red were genotyped in the second phase of the study, subsequent to the sequencing experiment. Middle panel shows $D'$ in the upper left and $r^2$ in the lower right. (**a**–**e**) Lower panels show results from association analysis of Leuven trios (**a**), Liège trios (**b**), Liège case-control cohort (**c**), Québec trios (**d**) and Toronto trios (**e**). *P* values for individual alleles are reported in a logarithmic scale on the *y* axis. Color spectrum represents strength of linkage disequilibrium and frequency of the associated alleles.

We next evaluated associations between genotype and gene expression to determine whether the above SNP associations reflected *cis*-acting regulatory effects on *NLRP3*. We first assessed the influence of the six SNPs (**Table 1**) on *NLRP3* mRNA expression from freshly isolated peripheral blood cells (PBCs) and from monocytes isolated from the peripheral blood, as *NLRP3* is primarily expressed in granulocytes and monocytes[9] (**Supplementary Methods**). A significant association was observed between *NLRP3* expression and rs4353135 genotypes in PBCs ($P = 0.00246$; **Fig. 2a**) and monocytes ($P = 0.0124$; **Fig. 2b**), with homozygosity for the risk allele being associated with the lowest level of *NLRP3* expression. Results for the five other Crohn's disease–associated SNPs are shown in **Supplementary Figure 1** and **Table 4** online.

Because *NLRP3* is involved in IL-1β processing, we therefore also evaluated whether these six SNPs influenced IL-1β production. We cultured monocytes in the presence or absence of crude lipopolysaccharide (LPS; **Supplementary Methods**), as its derivatives have been shown to stimulate *NLRP3* expression[10]. We then assessed IL-1β levels in culture supernatants. We observed a borderline-significant association between IL-1β levels and rs6672995 genotype under the unstimulated condition ($P = 0.0502$; **Fig. 2c**) and a significant association under the LPS-stimulated condition ($P = 0.00591$; **Fig. 2d**). In both

**Table 1 Replication studies and pooled analysis showing association between SNPs in the 5.3-kb region and risk of Crohn's disease**

| SNP | Associated allele | Control frequency[a] | Case frequency[b] | T:U ratio[c] | Control ratio[d] | Case ratio[d] | Odds ratio (95% CI)[e] | *P* value[e] |
|---|---|---|---|---|---|---|---|---|
| **Leuven (296 trios)** | | | | | | | | |
| rs4353135[f] | T | 0.65 | 0.73 | 118:82 | | | 1.44 (1.08–1.91) | 0.0107 |
| rs4266924[f] | A | 0.86 | 0.91 | 57:31 | | | 1.78 (1.14–2.79) | $7.63 \times 10^{-3}$ |
| rs55646866[g] | C | 0.86 | 0.93 | 57:26 | | | 2.19 (1.37–3.52) | $5.72 \times 10^{-4}$ |
| rs6672995[g] | G | 0.83 | 0.88 | 64:39 | | | 1.64 (1.09–2.48) | 0.0133 |
| ss107635144[g] | C | 0.84 | 0.91 | 64:36 | | | 1.78 (1.16–2.73) | $4.82 \times 10^{-3}$ |
| rs10733113[f] | G | 0.84 | 0.91 | 65:34 | | | 1.91 (1.24–2.95) | $1.68 \times 10^{-3}$ |
| **Liège trios (155 trios)** | | | | | | | | |
| rs4353135[f] | T | 0.62 | 0.73 | 33:20 | | | 1.65 (0.93–2.92) | 0.0727 |
| rs4266924[f] | A | 0.84 | 0.94 | 22:7 | | | 3.14 (1.36–7.26) | $4.31 \times 10^{-3}$ |
| rs55646866[g] | C | 0.85 | 0.95 | 20:6 | | | 3.33 (1.36–8.18) | $4.80 \times 10^{-3}$ |
| rs6672995[g] | G | 0.79 | 0.89 | 31:15 | | | 2.07 (1.11–3.84) | 0.0171 |
| ss107635144[g] | C | 0.80 | 0.90 | 27:13 | | | 2.08 (1.03–4.18) | 0.0253 |
| rs10733113[f] | G | 0.83 | 0.92 | 24:10 | | | 2.40 (1.16–4.96) | 0.0148 |
| **Liège CC (239 Crohn's disease and 107 controls)** | | | | | | | | |
| rs4353135[f] | T | 0.65 | 0.65 | | 39:44:11 | 92:95:27 | 1.01 (0.71–1.46) | 0.944 |
| rs4266924[f] | A | 0.79 | 0.90 | | 63:35:4 | 184:41:2 | 2.43 (1.53–3.85) | $1.62 \times 10^{-4}$ |
| rs55646866[g] | C | 0.80 | 0.91 | | 66:32:4 | 191:38:3 | 2.33 (1.47–3.69) | $4.32 \times 10^{-4}$ |
| rs6672995[g] | G | 0.74 | 0.86 | | 57:40:7 | 172:56:4 | 2.19 (1.46–3.30) | $1.87 \times 10^{-4}$ |
| ss107635144[g] | C | 0.75 | 0.86 | | 57:38:7 | 165:56:3 | 2.13 (1.41–3.23) | $3.94 \times 10^{-4}$ |
| rs10733113[f] | G | 0.76 | 0.88 | | 58:38:5 | 177:47:3 | 2.36 (1.52–3.65) | $1.20 \times 10^{-4}$ |
| **Québec (130 trios)[h]** | | | | | | | | |
| rs4353135[f] | T | 0.67 | 0.73 | 68:48 | | | 1.42 (1.00–2.01) | 0.0627 |
| rs4266924[f] | A | 0.85 | 0.91 | 43:26 | | | 1.65 (1.02–2.70) | 0.0400 |
| rs55646866[g] | C | 0.84 | 0.90 | 43:26 | | | 1.65 (1.02–2.70) | 0.0400 |
| rs6672995[g] | G | 0.78 | 0.85 | 53:32 | | | 1.66 (1.09–2.52) | 0.0220 |
| ss107635144[g] | C | 0.80 | 0.88 | 49:28 | | | 1.75 (1.12–2.73) | 0.0160 |
| rs10733113[f] | G | 0.79 | 0.89 | 53:28 | | | 1.89 (1.19–3.00) | $5.10 \times 10^{-3}$ |
| **Toronto (129 trios)[i]** | | | | | | | | |
| rs4353135[f] | T | 0.62 | 0.74 | 63:35 | | | 1.80 (1.17–2.77) | $4.40 \times 10^{-3}$ |
| rs4266924[f] | A | 0.82 | 0.92 | 39:16 | | | 2.50 (1.44–4.35) | $1.12 \times 10^{-3}$ |
| rs55646866[g] | C | 0.83 | 0.90 | 36:20 | | | 1.80 (1.07–3.04) | 0.0313 |
| rs6672995[g] | G | 0.77 | 0.85 | 46:28 | | | 1.64 (1.05–2.56) | 0.0355 |
| ss107635144[g] | C | 0.78 | 0.88 | 46:23 | | | 2.00 (1.22–3.27) | $5.18 \times 10^{-3}$ |
| rs10733113[f] | G | 0.77 | 0.91 | 50:17 | | | 3.00 (1.76–5.10) | $2.47 \times 10^{-5}$ |
| **Combined (710 trios, 239 Crohn's disease and 107 controls)[j]** | | | | | | | | |
| rs4353135[f] | T | 0.65 | 0.71 | 282:185 | 39:44:11 | 92:95:27 | 1.21 (1.05–1.39) | $8.36 \times 10^{-3}$ |
| rs4266924[f] | A | 0.84 | 0.91 | 161:80 | 63:35:4 | 184:41:2 | 1.69 (1.37–2.07) | $6.01 \times 10^{-7}$ |
| rs55646866[g] | C | 0.84 | 0.91 | 156:78 | 66:32:4 | 191:38:3 | 1.69 (1.38–2.08) | $7.20 \times 10^{-7}$ |
| rs6672995[g] | G | 0.79 | 0.87 | 194:114 | 57:40:7 | 172:56:4 | 1.53 (1.28–1.82) | $2.91 \times 10^{-6}$ |
| ss107635144[g] | C | 0.80 | 0.88 | 186:100 | 57:38:7 | 165:56:3 | 1.53 (1.27–1.84) | $8.50 \times 10^{-6}$ |
| rs10733113[f] | G | 0.80 | 0.90 | 192:89 | 58:38:5 | 177:47:3 | 1.78 (1.47–2.16) | $3.49 \times 10^{-9}$ |

[a]Estimated from untransmitted alleles or unrelated controls. [b]Estimated from transmitted alleles or unrelated cases. [c]Ratio of transmitted to untransmitted alleles from heterozygous parents. [d]Genotype distribution in cases and controls as x:y:z, where x = homozygous for the high-risk allele, y = heterozygous and z = homozygous for the second allele. [e]Odds ratios, confidence intervals (CI) and *P* values were computed from a likelihood test that assumes a multiplicative model for risk, as implemented in UNPHASED[28] (see Methods). Allele association was evaluated in the case-control cohort. All tests were two tailed. [f]SNPs associated with Crohn's disease in exploratory phase. [g]SNPs associated with Crohn's disease in second genotyping phase. [h]Includes 27 Crohn's disease trios of Jewish ancestry. [i]Includes 26 Crohn's disease trios of Jewish ancestry. [j]Includes 53 Crohn's disease trios of Jewish ancestry.

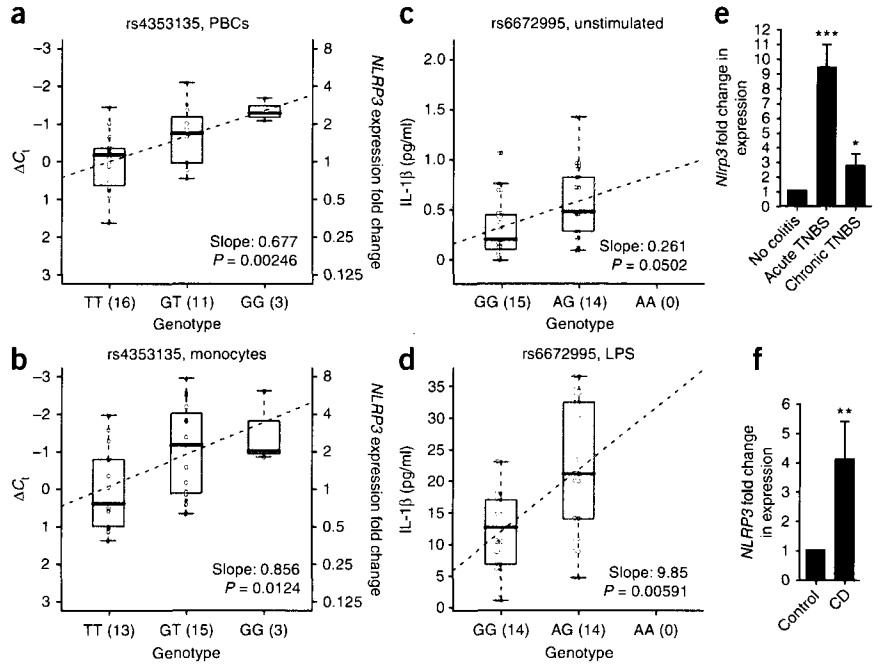



**Figure 2** *NLRP3* functional study results. (**a**,**b**) Linear regression analysis of *NLRP3* mRNA level versus rs4353135 genotype in DNA-RNA matched freshly isolated PBCs (**a**; $n = 30$) and monocytes (**b**; $n = 31$) obtained from healthy individuals. Genotypes of the six Crohn's disease (CD)-associated SNPs (**Table 1**) were obtained by sequencing. Mean threshold cycle ($C_t$) was calculated for each sample from three replicates and then used to calculate relative expression level ($\Delta C_t$), which is the difference between *NLRP3* $C_t$ and endogenous control 18S RNA $C_t$. Fold change in *NLRP3* expression was calculated using comparative $C_t$ method (see Methods), using as a reference the average $\Delta C_t$ of homozygosity for the risk allele. (**c**,**d**) Linear regression analysis of IL-1β production (pg/ml) versus rs6672995 genotype for unstimulated (**c**) and LPS-stimulated (**d**; 1.0 μg/ml) conditions after 3 h of incubation. $\Delta C_t$ (**a**,**b**) and IL-1β level (**c**,**d**) for each individual are shown in red; regression lines are shown as dashed lines (**a**–**d**). (**e**) Quantitative real-time PCR analysis of *Nlrp3* expression in colons of healthy mice ($n = 6$), mice with acute TNBS-induced colitis ($n = 12$) and mice with chronic TNBS-induced colitis ($n = 6$). (**f**) Quantitative real-time PCR analysis of *NLRP3* expression in colon specimens from healthy individuals ($n = 35$) and individuals with Crohn's disease ($n = 25$). Expression was normalized to 18S RNA expression; each bar represents mean fold change in *NLRP3* expression ± s.e.m. normalized to that of healthy colon specimens (**e**,**f**).



cases, homozygosity for the risk allele was associated with the lowest level of IL-1β. Results for the five other Crohn's disease–associated SNPs are shown in **Supplementary Figure 2** and **Table 5** online.

We also examined *Nlrp3* expression in colon tissues isolated from mice with trinitrobenzene sulfonic acid (TNBS)-induced colitis, a model that mimics Crohn's disease–like intestinal inflammation, and in biopsies from individuals with Crohn's disease. *Nlrp3* expression was significantly higher in both acute (fold change = 9.38 ± 1.58; $P < 0.0009$) and chronic (fold change = 2.70 ± 0.88; $P < 0.0152$) TNBS-induced colitis models than in colon tissues from control mice (**Fig. 2e**; see Methods and **Supplementary Methods**). *NLRP3* expression was also significantly higher in the ulcerated intestinal mucosa from human Crohn's disease samples (fold change = 4.08 ± 1.33; $P < 0.0028$) than in healthy controls (**Fig. 2f**; see Methods and **Supplementary Methods**).

*NLRP3* (chromosome 1q44) encodes cryopyrin, which is involved in the inflammasome signaling platform by regulating caspase-1 activity and IL-1β processing. The importance of cryopyrin in inflammation is highlighted by gain-of-function mutations within its NOD domain that are associated with three hereditary periodic fever syndromes: Muckle-Wells syndrome, familial cold autoinflammatory syndrome and neonatal-onset multisystem inflammatory disease[1,2]. Hyperproduction of IL-1β is thought to be a central event leading to symptoms in these three syndromes[1]. Consistent with these observations is the successful use of IL-1β targeted therapy for treating Muckle-Wells syndrome[11] and familial cold autoinflammatory syndrome[12].

In our study, contrary to the gain-of-function mutations described above leading to hyperproduction of IL-1β, we uncovered a regulatory region downstream of *NLRP3* that contributes to Crohn's disease susceptibility and is associated with hypoproduction of IL-1β and decreased *NLRP3* expression. Indeed, the risk allele of rs6672995, located in a predicted regulatory region[7], was associated with a decrease in LPS-induced IL-1β production, and the risk allele of rs4353135 was associated with a decrease in baseline *NLRP3* expression in two independent sample sets of healthy donors. It is noteworthy that these two SNPs were in weak linkage disequilibrium ($r^2 < 0.105$) in our combined sample set. Although the causal variant has not been conclusively shown and may still be unidentified, it is most likely to be in linkage disequilibrium with the tested variants. Nonetheless, we have shown that SNPs in the associated 5.3-kb region influence *NLRP3* at both the gene expression and functional levels, as indicated by altered *NLRP3* expression and IL-1β production. Notably, dysregulated IL-1β production has also been linked to Crohn's disease pathogenesis in which the three major *NOD2* (chromosome 16q12) mutations result in a loss-of-function phenotype, with decreased NF-κB activation in response to muramyl dipeptide stimulation and decreased IL-1β production in primary human mononuclear and dendritic cells from individuals with Crohn's disease[13–16]. Our results further support the recent Crohn's disease immunopathogenesis paradigm, which suggests that a defective innate immune response impairs clearance of luminal antigens and/or pathogens and leads to the development of chronic intestinal inflammation and Crohn's disease. *NLRP3* may thus have a role in the initiation phase of the disease, as indicated by our *in vitro* and expression experiments, as well as a role in perpetuating chronic inflammation through further activation of caspase-1 and processing of IL-1β, as indicated by the enhanced *NLRP3* expression in the Crohn's disease and chronic TNBS-induced colitis samples.

The *NLRP3* locus can be added to the list of several newly uncovered Crohn's disease loci at which the common allele has been reported to be the risk allele[17–22]. Although it is difficult to strictly distinguish between one allele being a susceptibility risk factor and the other being a protective one, estimating attributable fractions and prevented fractions in addition to odds ratios offers insight into how to interpret these associations with very common risk factors in the

context of complex diseases. If the SNPs with the strongest associations from **Table 1** are interpreted as risk factors, then the attributable fractions (that is, the reduction in prevalence if the risk factor were removed from the population) of the alleles fall in the range of 45–55%. These proportions of 'cases explained' (all other factors being ignored) are large, but not surprisingly so, as most members of the population are carriers of the risk factors. Conversely, if the minor alleles are interpreted as protective factors, then their prevented fractions (that is, the proportion by which the prevalence would increase if the protective factor were removed from the population) fall in the range of 10–14%, a range easier to interpret in the context of complex genetic diseases.

Several recent genome-wide association studies have identified new Crohn's disease susceptibility genes using the Illumina HumanHap300 Genotyping BeadChip[17–19] and the Affymetrix GeneChip Human Mapping 500K Array Set[21]. Although rs4353135 is present on the Human Mapping 500K Array Set, this SNP showed the weakest significance of all six SNPs from **Table 1**, was the only one not consistently replicated across all samples and was in weak linkage disequilibrium with the other SNPs (max $r^2 = 0.28$). **Supplementary Table 6** online shows the linkage disequilibrium between SNPs on the Illumina HumanHap300 and the Affymetrix GeneChip Human Mapping 500K arrays that were not genotyped in the present study and the SNPs from **Table 1** found in HapMap. With a maximum $r^2$ of only 0.16, these observations may explain why this region escaped detection in these genome-wide association studies.

A recent meta-analysis of three large genome-wide association studies of Crohn's disease reported that well-established associations with Crohn's disease account for ~20% of the genetic variance observed in Crohn's disease, suggesting that additional genetic contributions have yet to be discovered[22]. With the exception of variations within *NOD2* and *IL23R*, established susceptibility alleles have been reported to have relatively modest effects, with odds ratios ranging from 0.7 to 1.7 (ref. 22). Despite the modest contribution of *NLRP3* to the risk of Crohn's disease, our results strongly implicate *NLRP3*, a gene with an essential role in regulating the inflammasome, as a risk factor for Crohn's disease. Our results also suggest that a gene such as *NLRP3* that is associated with rare, severe autoinflammatory disorders can also be implicated in the susceptibility of more common inflammatory diseases such as Crohn's disease.



## METHODS

**Subjects.** Five sample sets from four different centers, totaling 710 Crohn's disease trios, 239 Crohn's disease cases and 107 controls, were assembled for this project (**Table 1**). All participants gave informed consent, and studies were approved by the Institutional Review Board of each institution that sent samples. A clinical subtype of Crohn's disease was assigned using standard clinical criteria[23,24], except for a few 'indeterminate colitis' cases that were excluded from the study (**Supplementary Methods**). Belgian subjects from Leuven University Hospital Gasthuisberg (Leuven) were used for the exploratory experiments ($n = 296$ trios). The replication cohorts consisted of Belgian subjects from Université de Liège and of two Canadian cohorts (Québec and Toronto). Samples from Liège were subdivided into a family-based cohort (155 Liège trios) and a case-control cohort (Liège case-control, 239 Crohn's disease and 107 controls). The Québec cohort ($n = 130$ trios) comprised subjects from multiple sites in the province of Québec and included 22 probands of Ashkenazi and 5 of Sephardic Jewish ancestry. The Toronto samples ($n = 129$ trios) were collected from multiple sites in Toronto and included 26 probands of Ashkenazi Jewish ancestry. All study participants were of European descent except for 15 probands from Toronto that were excluded from the analyses.

**Genotyping.** We first investigated a 67.8-kb region spanning 1q44 (243890897–243958709; NCBI build 35, hg17), including *NLRP3* (32.9 kb). SNP selection details are given in **Supplementary Methods**. Samples were genotyped using the SNPstream ultra-high-throughput genotyping system (Orchid Biosciences)[25] and Sequenom homogenous MassExtend (hME) assays (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry)[26]. Primers and probes are available in **Supplementary Table 1**. Analysis was restricted to SNPs passing quality filters, excluding SNPs with success rate < 95%, minor allele frequency < 5% or deviation from Hardy-Weinberg equilibrium ($P < 0.01$). Measures of linkage disequilibrium between SNPs and departures from Hardy-Weinberg equilibrium were computed using Haploview v4.0 (ref. 27). Families showing mendelian errors in 5% or more of the markers were excluded from the analysis (25 Crohn's disease families).

**Sequencing.** Primers were designed to have a $T_m$ of 60 °C using the Primer3 program (**Supplementary Table 2**). PCR reactions were done using Hot Start Taq polymerase (Qiagen) in an 8-μl final volume comprising 9 ng of DNA (one cycle of 96 °C for 10 min; 40 cycles of 95 °C for 30 s, 58 °C for 1 min and 72 °C for 35 s; and one cycle of 72 °C for 7 min). Sequencing was done on an ABI 3730 DNA sequencer (Applied Biosystems) according to standard protocols. Sequence traces were assembled and analyzed using the PolyPhred software package and were compared to annotated sequences from NCBI build 35, hg17.

**Statistical analysis.** Tests of association were done using the likelihood methods implemented in UNPHASED v3.0.10 (ref. 28), which can analyze samples of nuclear families, unrelated subjects or a combination of both. For nuclear families (cases and their parents), the likelihood is equivalent to the conditional likelihood models on which the transmission disequilibrium test is based[29]. For samples of unrelated cases and controls, the likelihood is equivalent to a logistic model, which allows estimations of risk effects in terms of odds ratios along with their confidence intervals. It also allows conditioning on the observed association at one marker, to test whether or not the observed significance at others can solely be explained by linkage disequilibrium. A permutation procedure is implemented to allow the calculation of significance levels that are corrected for the number of tests. Attributable and prevented fractions were calculated under the assumption that alleles had additive effects on the penetrance scale, that odds ratios from **Table 1** are good approximations for the relative risks and that allele frequencies in the controls are reasonable estimates of population frequencies.

**RNA extraction and quantitative real-time PCR.** Details on tissue collection and monocyte isolation are given in **Supplementary Methods**. Biopsies preserved in RNAlater (Qiagen) were homogenized, and total RNA was extracted using TRIzol (Invitrogen). Total RNA was extracted from monocytes using an RNeasy kit (Qiagen). Total RNA was extracted from PBCs using a PAXgene blood RNA kit (Qiagen) with RNase inhibitor, using off-column DNase I digestion and ethanol precipitation to improve RNA yield and quality. First-strand cDNA was synthesized from 1 μg of RNA template with a cDNA archive kit (Applied Biosystems), using MultiScribe reverse transcriptase and random primers. Quantitative real-time PCR assays (mouse *Nlrp3*, Mm00840904_m1; human *NLRP3*, Hs00918082_m1; 18S RNA, 4319413E; Applied Biosystems) were conducted using an ABI PRISM 7900 sequence detection system based on the 5′ nuclease assay[30] and quantified using Applied Biosystems' comparative threshold cycle ($C_t$) method. The Wilcoxon signed-rank test was used to evaluate tissue (mouse and human) expression differences. Associations between *NLRP3* expression or IL-1β level and the genotypes of the Crohn's disease–associated SNPs were assessed using linear regression. Analyses were done using GraphPad Software.

**DNA extraction.** Monocyte DNA was isolated from whole blood using a FlexiGene DNA kit (Qiagen). PBC DNA was isolated using a Gentra Autopure automated system (Qiagen) according to the manufacturer's protocol.

**Enzyme-linked immunosorbent assay.** IL-1β was quantified in monocyte culture supernatants using a human IL-1β DuoSet ELISA kit (R&D Systems) according to the manufacturer's protocol.

**URLs.** ESPERR software, http://www.bx.psu.edu/projects/esperr.

**Accession codes.** GenBank: *NLRP3*, AF054176; *OR2B11*, NM_001004492. OMIM: *NLRP3*, 606416; *OR2B11*, 605956. SNP data have been submitted to

NCBI dbSNP under the numbers ss107635120, ss107635122, ss107635124, ss107635126, ss107635128, ss107635130, ss107635132, ss107635133, ss107635136, ss107635138, ss107635140, ss107635142, ss107635144, ss107635146.

*Note: Supplementary information is available on the Nature Genetics website.*


## ACKNOWLEDGMENTS

We thank the study participants and their families for their collaboration; A. Montpetit and the genotyping and sequencing platforms of the McGill University and Genome Québec Innovation Centre for their technical assistance; T. McKenzie, G. Bonventi and C. Landolt-Maticorena for their help with acquisition and preparation of the PBCs; F. Martinon and J. Tschopp for their technical assistance in designing the functional studies; and T. Kwan for helpful discussion. This research was supported by grants to D.F. from the Canada Foundation for Innovation, the Canada Research Chair program, the Research Institute of the McGill University Health Centre and the Crohn's & Colitis Foundation of Canada (CCFC). A.-C.V. is supported by a PhD training award from the Fonds de la recherche en santé du Québec. M.S.S. is funded by the National Institute of Diabetes and Digestive and Kidney Diseases (DK624230), the CCFC and the Mount Sinai Hospital Gale and Graham Wright Research Chair in Digestive Diseases. J.D.R. is funded by grants from the National Institutes of Allergy and Infectious Diseases (AI065687 and AI067152), the National Institute of Diabetes and Digestive and Kidney Diseases (DK064869 and DK062432) and the Crohn's and Colitis Foundation of America (SRA512). E.L. is a senior research associate at the Belgium Foundation for Scientific Research (FNRS). S.V. is a clinical investigator for the Funds for Scientific Research of Flanders. T.J.H. is the recipient of an Investigator Award from the Ontario Institute for Cancer Research through funding from the Government of Ontario's Ministry of Research and Innovation. D.F. is a senior clinical scientist at the Belgium FNRS. None of the funding agencies had any role in the design and conduct of the study; in the collection, analysis and interpretation of the data; or in the preparation, review or approval of the manuscript.


## AUTHOR CONTRIBUTIONS

A.-C.V., G.F. and D.F. conceived and designed the experiments. A.-C.V., G.F., C.C. and N.B. did the experiments. A.-C.V. and M.L. analyzed the data. E.L., M.S.S., C.L., J.B., A.B., D.G., A.C., D.L., P.R.F., J.E.W., M.S., P.R., J.D.R., S.V., T.J.H. and D.F. provided study samples. A.-C.V., M.L. and D.F. wrote the paper, with contributions from all authors.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/




1. Agostini, L. *et al.* NALP3 forms an IL-1beta-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. *Immunity* **20**, 319–325 (2004).
2. Mariathasan, S. & Monack, D.M. Inflammasome adaptors and sensors: intracellular regulators of infection and inflammation. *Nat. Rev. Immunol.* **7**, 31–40 (2007).
3. Podolsky, D.K. Inflammatory bowel disease. *N. Engl. J. Med.* **347**, 417–429 (2002).
4. Ting, J.P., Kastner, D.L. & Hoffman, H.M. CATERPILLERs, pyrin and hereditary immunological disorders. *Nat. Rev. Immunol.* **6**, 183–195 (2006).
5. Pétrilli, V., Dostert, C., Muruve, D.A. & Tschopp, J. The inflammasome: a danger sensing complex triggering innate immunity. *Curr. Opin. Immunol.* **19**, 615–622 (2007).
6. Martinon, F. & Tschopp, J. Inflammatory caspases and inflammasomes: master switches of inflammation. *Cell Death Differ.* **14**, 10–22 (2007).
7. Taylor, J. *et al.* ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. *Genome Res.* **16**, 1596–1604 (2006).
8. de Bakker, P.I. *et al.* Efficiency and power in genetic association studies. *Nat. Genet.* **37**, 1217–1223 (2005).
9. Kummer, J.A. *et al.* Inflammasome components NALP 1 and 3 show distinct but separate expression profiles in human tissues suggesting a site-specific role in the inflammatory response. *J. Histochem. Cytochem.* **55**, 443–452 (2007).
10. Martinon, F., Agostini, L., Meylan, E. & Tschopp, J. Identification of bacterial muramyl dipeptide as activator of the NALP3/cryopyrin inflammasome. *Curr. Biol.* **14**, 1929–1934 (2004).
11. Hawkins, P.N., Lachmann, H.J. & McDermott, M.F. Interleukin-1-receptor antagonist in the Muckle-Wells syndrome. *N. Engl. J. Med.* **348**, 2583–2584 (2003).
12. Hoffman, H.M. *et al.* Prevention of cold-associated acute inflammation in familial cold autoinflammatory syndrome by interleukin-1 receptor antagonist. *Lancet* **364**, 1779–1785 (2004).
13. Li, J. *et al.* Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. *Hum. Mol. Genet.* **13**, 1715–1725 (2004).
14. Van Heel, D.A. *et al.* Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease. *Lancet* **365**, 1794–1796 (2005).
15. Kramer, M., Netea, M.G., de Jong, D.J., Kullberg, B.J. & Adema, G.J. Impaired dendritic cell function in Crohn's disease patients with NOD2 3020insC mutation. *J. Leukoc. Biol.* **79**, 860–866 (2006).
16. van Beelen, A.J. *et al.* Stimulation of the intracellular bacterial sensor NOD2 programs dendritic cells to promote interleukin-17 production in human memory T cells. *Immunity* **27**, 660–669 (2007).
17. Duerr, R.H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
18. Rioux, J.D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* **39**, 596–604 (2007).
19. 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.* **3**, e58 (2007).
20. Parkes, M. *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.* **39**, 830–832 (2007).
21. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
22. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
23. Lennard-Jones, J.E. Classification of inflammatory bowel disease. *Scand. J. Gastroenterol. Suppl.* **170**, 2–6 (1989).
24. 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.* (Suppl A), **19**, 5–36 (2005).
25. Bell, P.A. *et al.* SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques* Suppl 70–72, 74, 76–77 (2002).
26. van den Boom, D. & Ehrich, M. Discovery and identification of sequence polymorphisms and mutations with MALDI-TOF MS. *Methods Mol. Biol.* **366**, 287–306 (2007).
27. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
28. Dudbridge, F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum. Hered.* **66**, 87–98 (2008).
29. Spielman, R.S., McGinnis, R.E. & Ewens, W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516 (1993).
30. Holland, P.M., Abramson, R.D., Watson, R. & Gelfand, D.H. Detection of specific polymerase chain reaction product by utilizing the 5′→3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* **88**, 7276–7280 (1991).