FUNCTIONAL VARIATIONS OF ORGANIC CATION TRANSPORTERS ASSOCIATED TO INFLAMMATORY BOWEL DISEASE

by

Alejandra Serrano León

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Human Nutritional Sciences

University of Manitoba

Winnipeg

Copyright © 2013 by Alejandra Serrano León

ABSTRACT

Organic cation transporters are involved in the absorption, distribution and excretion of therapeutic drugs and food compounds. Polymorphisms in *SLC22A4*, *SLC22A23* and *IBD5* locus have been associated with pathogenesis of inflammatory bowel disease (IBD). We sought to investigate the association of some common and rare SNPs to IBD risk in a Canadian population, subclone and express the two main isoforms of human *SLC22A23* gene, and determine the localization in the cell. DNA samples from 160 patients with Crohn´s disease (CD), 149 patients with ulcerative colitis (UC) and 142 healthy controls were genotyped by PCR-RFLP analysis or TaqMan system for polymorphism in *SLC22A4*, *SLC22A23* and *IBD5* locus. Human *SLC22A23* isoforms A and B were subcloned from placenta cDNA with N-terminal mCherry and tGFP fluorescent proteins, respectively. Gateway® recombination technology and pcDNA™/V5-DEST expression vector were used to transform and express both isoforms in HEK 293 cell line. Fluorescent microscopy was used to determine the cell localization of *SLC22A23* isoforms. Polymorphisms in the IBD5 locus rs17622208-AA genotype and rs11739135-CC genotype increase the risk of CD (OR=2.26 and OR=7.84, respectively). Moreover, carriers of *SLC22A23* polymorphisms rs4959235-TT genotype and rs9503518-GG genotype increase dramatically the risk of UC (OR= 11.42 and OR= 16.71, respectively). We confirm that *SLC22A23* protein is localized in the cell membrane. The results confirm the importance of genetic variations in organic cation transporters in the pathogenesis of IBD and make them potential therapeutic targets.

ACKNOWLEDGMENTS

I would like to express my gratefulness and admiration to my supervisor Dr Peter Eck, for the confidence, inspiration, invaluable knowledge and support throughout my Master´s program. Without him my time at the University of Manitoba and the completion of the thesis would not have been possible.

I extend my appreciation to my committee members Dr Harold Aukema and Dr Tedros Bezabeh for their support and contribution in this project.

My most sincere appreciation to the Faculty of Graduate Studies and the Government of Manitoba for their financial assistance.

My greatest thank you to my dear friend Sandra Castillo San Juan for the unconditional support and for making this experience memorable.

Finally, I would like to thank my parents Alma and Alejandro, my sister Gabriela and my dear Manuel for motivating and encouraging me to succeed and for all the support and love during this dream.

TABLE OF CONTENTS

LIST OF TABLES

Manuscript 1

Manuscript 2

LIST OF FIGURES

Manuscript 1

Manuscript 2

Manuscript 3

CHAPTER 1. INTRODUCTION

Background and Rationale

Many common chronic diseases are caused by multiple genes interacting with various environmental factors, one of them the diet. It is well known that in nutritional practice of one recommendation fits all can be improved when considering an individual's genetically determined response to a specific dietary component. The novel field of Nutrigenomics and Nutrigenetics promise to deliver individualized gene specific dietary recommendations.

Membrane transporters are of great physiological importance as they are the major determinants for the absorption, distribution, and elimination of a large number of dietary components. Therefore they are prime candidates to study how genetic variations alter nutrient pharmacokinetics to increase susceptibility to diseases. Once associated to a disease the genetic variations have to be characterized with respect to how they change the protein's function.

Crohn's disease is an inflammatory autoimmune syndrome of the digestive system with the highest incidence in Canada and particularly in the province of Manitoba[1,2]. Several genetic studies have shown associations between mutations in organic cation transporters *SLC22A4* and *SLC22A5* located in what has been called *IBD5* locus[3,4]. SLC22A5 transports mainly carnitine[5] by contrast, the molecular and physiologic function of SLC22A4 is less understood, but is proposed to be the main transporter of the antioxidant ergothioneine[6].

Recently, the orphan and still uncharacterized organic cation transporter *SLC22A23* was associated with Crohn's disease in a genome-wide association study[7]. Several genetic studies show an association of *SLC22A4* and *SLC22A5* single nucleotide polymorphism (SNPs) or haplotypes to Crohn's disease. However, studies in cohorts of different regional origin and ethnic backgrounds show substantial discrepancies in disease associations. To date, no replication study has confirmed the involvement of *SLC22A23* in any IBD cohort.

To contribute to the understanding if *SLC22A23*, *SLC22A4*, and *SLC22A5* are involved in the disease susceptibility in a Manitoban Caucasian population based cohort we genotyped selected tag SNP for an association study.

To understand the molecular mechanism of *SLC22A23*´s disease contribution we characterized the gene and its transcripts variants and subcloned the major isoform to determine cellular localization on the plasma membrane.

In the long term we hope to be able to determine disease mechanism, which will enable the development of population specific dietary intervention strategies.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

The following review addresses the up-to-date knowledge regarding the association of polymorphism in *IBD5* locus, *SLC22A4*, *SLC22A5* and *SLC22A23* genes to inflammatory bowel disease (IBD) and the functional evidence of their involvement in the disease.

2.2 Inflammatory Bowel Disease and Classification

Inflammatory Bowel Disease (IBD) is considered a chronic inflammatory disorder of the gastrointestinal tract, characterized by progressive damage of the bowel with periods of relapse and remission of clinical symptoms. Hitherto, the mechanisms or factors responsible for triggering the disease are not yet identified. Rather, it has been suggested a complex interplay of environmental and genetic factors along with immune dysregulation[8].

IBD is classified based on clinical manifestations, radiologic and histological parameters into three principal disease groups[9]:

- **Crohn´s disease (CD):** is characterized by inflammation at any level of the gastrointestinal tract, from mouth to anus, it is also discontinuous and transmural, thus affecting from mucosa to serosa. Associated with strictures, fistulas and abscesses.
- **Ulcerative Colitis (UC):** is characterized by exclusive inflammation of mucosa at the colon level, in a continuous way.
- **Indeterminate Colitis (IC):** Around 10 to 15% of IBD patients are classified as IC, due to a difficulty in making a differentiation.

Some clinical manifestations are shared among the three groups, such as abdominal pain, diarrhea, urgency bloody stools, weight loss and fever. However, the phenotype will vary among patients leading to a great inter-variability among each category[9]

The large variability among IBD patients, led to the development of IBD subclassification guidelines, which allow selecting the best therapeutic approach, and predicting disease prognosis, as well as a better understanding of the pathophysiological mechanisms involved in IBD phenotypes[10]. The Montreal classification is the most recent classification of CD and UC. It was established in 2005, but to insufficient knowledge at the time does not include molecular markers such as genetics and serology markers.

The variables included in the Montreal classification for CD are age at onset of disease, anatomic location and disease behavior. For UC it includes age at onset, disease extent and severity. In 2009, it was recognized that some pediatric phenotypes are very dynamic and the Paris classification was established for clinical practice in pediatric IBD patients [10, 11] (Table 1).

Table 1. Montreal and Paris Classification of IBD [10,12]

	Montreal (2005)	Paris (2009)
Age at Onset	A1- Below 16 years A2- between 17 and 40 years A3- above 40 years	A1a- from 0 to <10 years A1b- from 10 to <17 years A2- between 17 and 40 years A3- above 40 years
Location	L1- Ileal L2- Colonic L3- Ileocolonic L4- Upper disease *	L1- Ileal L2- Colonic L3- Ileocolonic L4a- Upper disease proximal to ligament of Treitz * L4b- Upper disease distal to ligament of Treitz *
Behavior	B1- non-stricturing, non-penetrating B2- stricturing B3- penetrating p- perianal disease **	B1- non-stricturing, non-penetrating B2- stricturing B3- penetrating B2B3- structuring and penetrating p- perianal disease **
Growth	Not included	G_0- No evidence of growth failure G_1- Growth failure
***L4 is a modifier of either L1-L3 ** p is a modifier of either B1-B3**		

2.3 Epidemiology

Prevalence of IBD in the Canadian population is about 1 in 180, approximately 170,000 Canadians have IBD, representing one of the highest incidence rates in the world. This is in accordance to the observation that northern countries present higher rates than southern counties. The overall incidence and prevalence of CD across Canada was estimated to be 16.3/100,000 and 279/100,000, respectively, and the overall incidence and prevalence for UC was estimated to be 11.8/100,000 and 211/100,000, respectively[1]. In 2006, it was

reported from a population base study that Canada, specifically the province of Manitoba, present the highest rates of incidence and prevalence of CD for the years 1990-2001 of 14.8/100,000 and 222.2/100,00, respectively. The incidence and prevalence of UC for the same period is 14.6/100,000 and 197.9/100,000, respectively[2].

2.4 Genetic Factors

Significant variations in the incidence and prevalence rates of IBD between different populations suggest a role of genetic factors which confer susceptibility to the disease. High prevalence has been reported for Caucasians and Jewish ethnicities, as well as higher rates among monozygotic twins compared to dizygotic twins. In addition, first-degree relatives of a patient have 4 to 20 times greater risk of developing the disease[13].

The development of genome-wide association studies (GWAS) have allowed a better understanding of the genetic factors associated to IBD. The International IBD Genetic Consortium (IIBDGC) have identified 99 IBD susceptibility loci, of which 71 are associated to CD, 47 to UC and 28 are shared[14].

2.5 The role of IBD5 locus

The IBD5 locus is a genomic region localized in chromosome 5q31. It was first associated to CD by Rioux et al[15] through a linkage analysis of Canadian affected sibling pair families. Due to the immunological dysregulation in IBD, this genomic region was of paramount importance because it contains interleukin-4 (IL4), IL13, IL5 and interferon regulatory factor-1 (IRF1), but it also contains two organic cation/carnitine transporters *SLC22A4* and *SLC22A5*. Hence, additional markers were used and a 500kb region was delimited, which was conferring the risk. However, no single causative variation has been

identified due to the high linkage disequilibrium in the region. The resequencing of eight individuals allowed the identification of candidate SNPs, which were genotyped in CD trios and led to the identification of the risk haplotype of approximately 250 kb[16].

Later, Peltekova et al[17] identified two functional polymorphisms within *SLC22A4* and *SLC22A5*. A nonsynonymous polymorphism in *SLC22A4* (rs1050152_C1672T) causes a conversion of leucine 503 to phenylalanine (L503F) in the protein known as OCTN1. It was reported that fibroblast expressing the T allele had 2.7 times lower carnitine uptake compared to the C allele. Also, the uptake of the xenobiotic TEA was increased in the isoform carrying the T allele.

Another polymorphism associated to CD in *SLC22A5* (rs2631367_G-207C) lies in the 5´UTR in a heat shock transcription factor (HSF)-binding element (HSE). Promoter activity was tested and fibroblast expressing the C allele showed 2.3 times lower promoter activity compared to the G allele. Therefore, it was suggested that this SNP is important to regulate the transcription of *SLC22A5*.

Peltekova et al[17] suggested that both polymorphisms are associated to CD, independently from the IBD5 locus.

A replication study within a Canadian IBD cohort showed that polymorphisms in *SLC22A4* (rs1050152-T) and *SLC22A5* (rs2631367-C) are associated with ileal CD but not with UC. Moreover, these authors suggest the TC haplotype as causative, since it was independent from IBD5 haplotype, however, they used the same tagging SNP (IGR2078) as Peltekova et al. for the IBD5 locus which is outside the distinct the 250kb IBD5 risk haplotype[18].

In contrast, Török et al[19] found that *SLC22A4/5* variants in a German population are associated with CD only in the presence of the risk genotype IGR2078 in the IBD5 locus. They also found the *SLC22A4/5* TC haplotype was associated with colonic involvement and earlier age at disease onset, considering 22 years as a median age.

Similar results were found in a Scottish IBD cohort, where *SLC22A4* (rs1050152-T) and *SLC22A5* (rs2631367-C) genotypes were associated to CD only in the presence of other tagging polymorphisms of the *IBD5* locus (rs12521868-T, rs11739135-C and rs17622208-A). Interestingly, they report an association of the rs11739135-C and *SLC22A4/5* TC haplotype with the presence of structuring and penetrating disease and the need for surgery, which suggest that these polymorphisms might be implicated in the development of more severe phenotypes[20]. This was in accordance with the association of the same polymorphisms in IBD5 locus and *SLC22A4/5* TC haplotype to CD and UC in a pediatric Scottish population. They also reported association with low weight and low Body Mass Index (BMI) at diagnosis, which can be used as severe phenotype predictor[21].

Various haplotype combinations have been implicated to indicate elevated IBD risk. In a Spanish cohort *SLC22A4* rs1050152-TT and *SLC22A5* rs2631367-CC genotypes increase CD risk only in the presence of the IBD5 wild-type IGR3081-TT genotype[22]. The IGR3081-GG genotype was strongly associated with CD only when the rs1050152-TT and rs2631367-CC genotypes were absent[22]. A Caucasian Czech population study concludes that IBD5 locus is associated to the disease and rs6596075-CC genotype double the risk of penetrating CD, independently from the risk haplotype[23].

In 2007, using a much larger Caucasian population cohort which included IBD offspring and parents, Silverbeg et al., reported strong association of the IBD5 rs12521868-TT genotype to CD. This SNP is located close to the *SLC22A5* and *IRF1* genes. They also reported association of rs1050152-TT genotype in *SLC22A4* and rs11739135-CC genotype with CD and UC. In this study, they suggest that *SLC22A5* rs2631367-CC is not causative, since due to the strong linkage disequilibrium the effects of the *SLC22A4/5* TC haplotype could not be distinguished from others[24].

The association of rs12521868-TT and rs11739135-CC genotype to CD was confirmed in a Hungarian cohort (OR=1.68 and OR= 1.73, respectively). However, both polymorphisms were not associated to UC[25].

Interestingly, a study of the Slovenian IBD population reported lower expression of *SLC22A5* and *SLC22A4* in peripheral blood lymphocytes in carriers of the rs1050152-TT and rs2631367-CC genotypes. A trend towards a lower expression of both genes in inflamed tissue compared to noninflamed tissue in carriers of the IBD risk genotypes was observed[26].

In 2011 a meta-analysis was performed assessing the association of SNPs in the *IBD5* locus and *SLC22A4* and *SLC22A5* to IBD. The most consistent tagSNPs which replicated IBD associations were included in the analysis, the IBD5 locus (IGR2096, IGR2198 and IGR2230) and the functional polymorphism of SLC22A4 (rs1050152) and SLC22A5 (rs2631367). The results of the meta-analysis suggest that IBD5 increases the risk of CD in Caucasians adults and children, and increases the risk of UC in a recessive manner in adults from the same ethnicity[3].

Another meta-analysis that includes several association studies of the IBD5 locus in diverse Caucasian populations, confirmed that polymorphism in *SLC22A4* and *SLC22A5* increase the susceptibility to CD in Caucasian cohorts (OR=1.34, p=0.000; OR=1.23, p=0.018, respectively) but they are not associated in East Asian population, who present only the ancestral alleles[4].

In contrast to the common disease common genotype model of inheritance, the rare allele model postulates that genes containing common variants might also contain rare variants with dominant effect, increasing the disease-susceptibility more than two fold. Urban et al., revealed that the rare variants rs11568500 and rs11568510 within the *SLC22A4* gene are responsible for 50% reduction and completely loss of function of the protein, respectively[27]. Therefore, including these into genetic analysis of IBD patients will help elucidating if these rare variations have a role in disease susceptibility.

Overall, the causal polymorphisms in the IBD5 locus remain undetermined, but mounting evidence indicates a role of SNPs in *SLC22A4* and *SLC22A5* in disease susceptibility.

2.6 Organic Cation/Carnitine Transporters SLC22A4 and SLC22A5

The organic cation carnitine transporters are encoding by the genes SLC22A4 and SLC22A5. They are members of the solute carrier family SLC22, which belongs to the major facilitator superfamily. The functional proteins encoded by both genes are called OCTN1 and OCTN2, respectively. The amino acid sequence among them is 75.8% similar[5]. They transport organic cations, and are able to recognized carnitine (beta-hydroxygamma-trimethylaminobutyric acid)[28].

2.6.1 The role of SLC22A4 in inflammatory bowel disease

It was in 1997 when SLC22A4 was first characterized sharing 30% of the amino acid sequence with other organic transporters and a nucleotide-binding site motif that is unique to this transporter. It is expressed in bone marrow $CD71^{+}$, $CD34^{+}$ cells, kidney, small intestine, fetal liver, lung, trachea, cerebellum, placenta, mammary gland, prostate and is less expressed in skeletal muscle, heart, skin. When expressed in HEK293 cells, it transported the xenobiotic tetraethylammonium (TEA) with a lower activity at acidic pH, it was therefore considered as a H^{+}/organic cation antiporter[29]. It is a multispecific transporter, transporting gabapentin, pyrilamine, quinidine, verapamil, hydroxyproline, stachydrine, and with a low affinity L-carnitine and acetylcarnitine, but it showed high affinity to ergothioneine[6,29]. Although the physiological role of SLC22A4 is not fully determined, today it is considered as the principal transporter of ergothioneine (ET) in a Na^{+} dependent manner, by what is known as ergothioneine transporter (ETT)[6].

Ergothionine (ET, 2-mercaptohistidine trimethylbetaine) is a water soluble thione derivative of the histidine amino acid. It is a very stable antioxidant and does not promote the generation of hydroxyl radicals. It is produced by fungi and mycobacteria in the soil through where it is delivered to the plants. Therefore, humans obtain it through the intake of mushrooms and cyanobacteria such as spirulina, which are the best dietary sources of ET. There is a correlation between the *SLC22A4* expression and the accumulation of ET, thus acting as biomarker for the ET activity[30].

The functional effect of the IBD associated non-synonymous polymorphism rs1050152-T in *SLC22A4* is not yet fully understood. Peltekova et al., report a reduction in the uptake of carnitine when SLC22A4 expressing fibroblast carry the rs1050152-T allele, compared to the wild-type[17]. Contrary, when ET is considered as the main substrate, carriers of the rs1050152-T allele have higher uptake of ET into the cell, compared to the wild-type rs1050152-C allele. Therefore, it is sometimes suggested that high concentration of ergothioneine are involved in the CD development[31]. In accordance with these hypothesis, either CD patients or healthy controls had higher levels of ergothioneine in non-inflamed mucosal biopsies when carrying rs1050152-TT genotype compared to the wild-type. In inflamed mucosal biopsies from CD patients, carriers of the rs1050152-TT genotype had two fold the amount of ergothioneine compared to non-inflamed mucosa[32].

The antioxidant activity of ET might be important for IBD pathophysiology, however it must be stressed that there are no reports whether ET is a more potent antioxidant than glutathione or ascorbate, especially if they are at higher intracellular concentrations[33].

Moreover, its presence in a restricted range of foods does not support an essential role of ET, especially in the development of IBD.

With the aim to understand the physiological role of *SLC22A4*, a knockout mice model was created (slc22a4$^{-/-}$). Apparently, the slc22a4$^{-/-}$ mice did not show any phenotypic abnormality compared to the wild-type mice. But, further metabolome comparison revealed that ergothioneine was absent in erythrocytes, heart, liver, small intestine and kidney of the slc22a4$^{-/-}$. At the same time, slc22a4$^{-/-}$ mice were able to absorb ergothioneine after oral administration, which lead to the thinking of the existence of additional ergothioneine transporters. After ischemia and reperfusion to create an inflammatory model, a higher lethality and loss of villus structures in small intestine of the slc22a4$^{-/-}$ compared to the wild-type[34] mice was observed.

The same group reported that CD patients had lower blood concentrations of ergothioneine than UC and also healthy people. They even propose it as a diagnostic tool to classify both conditions[34].

However, mushrooms are the best dietary source of ergothioneine and Petermann et al, reported that New Zealand CD patients had adverse effects when consuming maize and mushrooms. Moreover, when stratified by *SLC22A4* rs1050152-TT genotype, CD carriers of that genotype show higher sensitivity to mushroom intolerance. It is important to consider that it was not an intervention study and the ergothioneine blood levels or SLC22A4 expression was not taken into account[35].

Despite that it has been suggested ergothioneine as the main substrate for *SLC22A4*; there are many drawbacks as to think that a more physiological substrate remains

uncharacterized. In this regard, Pochini et al., suggest that SLC22A4 mediates the uptake and efflux of acetylcholine, this was tested in a model of reconstitution in liposomes. In addition, they reported that rs1050152-T polymorphism reduced the export of acetylcholine. They hypothesize that rs1050152-T polymorphism compromises the barrier integrity, as in IBD, through defective export of acetylcholine, which has a role in maintaining epithelial integrity and cell-cell interaction[36].

Further functional studies are needed to fully determine the truly physiological role and substrate of SLC22A4 and determine if the rs1050152 polymorphism is the real causative variation in IBD or a more complex interaction of several genetic variations influence the mechanism by which *SLC22A4* is associated to IBD.

2.6.2 The role of *SLC22A5* in inflammatory bowel disease

In 1998, *SLC22A5* was first cloned and expressed in HEK293 cells where it showed high transport for carnitine in a sodium dependent manner. The gene is expressed in a wide range of tissues such as kidney, intestine, skeletal muscle, liver, lung, brain, placenta, pancreas, heart, prostate, thyroid and spinal cord[5]. Mutations in *SLC22A5* lead to systemic carnitine deficiency (SCD)[37].

A single nucleotide polymorphism (rs2631367-C) in the *SLC22A5* gene has been associated with IBD. Peltekova et al., first reported that rs2631367-C causes a down-regulation of the promoter activity, thus decreasing the expression of the transporter[17].

Carnitine is an essential metabolite that regulates the transport of long-chain fatty acids into the mitochondria for the generation of ATP by β-oxidation. This is particularly important for the enterocytes to maintain normal gut morphology and function. The

importance of carnitine in the intestinal inflammation is supported by Shekhawat et al., using a Slc22a5$^{-/-}$ knockout mice model of systemic carnitine deficiency. In this model, the levels of carnitine in the gut are reduced by 90% and it presents spontaneous perforations, micro-abscess, necrotic villi which in turn lead to apoptosis of gut epithelial cells and gut atrophy[38].

Sonne et al. reported that Slc22a5$^{-/-}$ mice present alterations in villus architecture, lymphocytic infiltration, there is a reduction in the mucin producing globet cells and increase in macrophages in the ileum. The thymus and spleen of this mice model was much smaller and pale than the wild-type model. It was also reported a reduction in the total count of lymphocyte produced by thymus, spleen and lymph nodes. Moreover, lymphocytes had a higher apoptosis rate. This process makes the gut barrier more susceptible to the entry of pathogens which initiate the inflammatory process[39].

In a different IBD mouse model, the trinitrobenzene sulfonic acid (TNBS)-induced colitis, carnitine was supplemented to the mice and it was responsible for the reduction in inflammation and histological damage. Carnitine also reduces the mRNA expression of serum levels of pro-inflammatory cytokines IL-1β and IL-6[40].

Moreover, the supplementation with propionyl-L-carnitine (1g/day) to UC patients induces improvements in the clinical and endoscopic response, as well as the induction of remission at higher frequency in patients taking the supplement than in patients under placebo[41].

The mechanism by which *SLC22A5* is implicated in the development of IBD is better comprehended compared to the mechanism of *SLC22A4*, however; still remains the

challenge to identify the true causal variations associated to intestinal inflammation, particularly in a high risk population such as the Manitoba population.

2.6.3 The role of SLC22A23 in inflammatory bowel disease

The human *SLC22A23* gene was first identified by Jacobsson et al, in 2007 according to the similar genetic sequence it is in the SLC22 family. However, little is known about its physiological function because the genomic organization, localization and genetic variations have not been described. The *SLC22A23* gene is expressed in a wide range of tissues such as bladder, blood, bone, brain, cervix, eye, heart, intestine, kidney, liver, lung, muscle, pancreas, placenta, and trachea. It even has been found in several tumor tissues like colorectal and gastrointestinal tumors[42].

Human *SLC22A23* has never been cloned or expressed to study its localization and function. However, Bennett et al., investigated the rat *slc22a23* ortholog. They reported that the expressed protein is located in the cell membrane. Further transport analyses reveal that *slc22a23* does not transport carnitine or TEA, typical substrates for OCTN members. A difference was shown in the cellular accumulation of histamine and ascorbate, compared to the vector alone, slc22a23 expressing cells had lower uptake of both compounds, suggesting the potential export function of this transporter[43].

A genome-wide association study, implicated the *SLC22A23* gene to IBD through the association of rs17309827-T allele[7]. But, no independent replication study had been conducted to confirm the association.

CHAPTER 3. STUDY OBJECTIVES

The objectives of this study are:

1. To determine the association with IBD of some common and rare single nucleotide polymorphism (SNPs) in the *IBD5* locus, specifically the *SLC22A4* and *SLC22A5* genes in Caucasian Manitoban population based cohort.

2. To determine the association of some common and rare single nucleotide polymorphisms (SNPs) in *SLC22A23* to IBD risk in a Manitoba cohort.

3. To characterize the structure of the human *SLC22A23* gene as well as the splice variants.

4. To subclone and express the human *SLC22A23* gene in mammalian cell culture.

5. To determine the localization of the SLC22A23 expressed protein within the cell.

CHAPTER 4. Manuscript 1: Variations in the *IBD5* locus confer the risk of Inflammatory Bowel Disease in a Manitoban Caucasian Cohort

Alejandra Serrano León[1, 2], Natalia Yurkova[2], Charles N. Bernstein[3], Hani El-Gabalawy[4], Peter Eck[1, 2]

[1]Human Nutritional Sciences, University of Manitoba, [2] The Richardson Centre for Functional Foods and Nutraceuticals, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, [3]University of Manitoba IBD Clinical and Research Centre, [4]Department of Internal Medicine, University of Manitoba, Winnipeg, Canada, R3E 3P4.

4.1 ABSTRACT

Background: Crohn´s disease (CD) and ulcerative colitis (UC) are two distinct manifestations of inflammatory bowel disease (IBD). Polymorphisms in *SLC22A4*, *SLC22A5* genes, both linked within the IBD5 locus, have been associated with pathogenesis of IBD. **Objective:** This study sought to investigate the association of common and rare variations within the *SLC22A4* and *SLC22A5* genes in a Manitoban Caucasian IBD cohort. **Design:** DNA samples from 160 CD patients, 149 UC patients and 142 age and gender matched healthy controls were genotyped by PCR-RFLP analysis. **Results:** The *SLC22A5* genotypes rs11739135-CC and rs17622208-AA increased the susceptibility to CD (OR=7.84, 95% CI 2.84-21.6, p=0.000; OR=2.26, 95% CI 1.14-4.44, p=0.019, respectively). Moreover, rs11739135-CC genotype was associated with UC (OR=4.18, 95% CI 1.48-11.78, p=0.007). The rare polymorphisms rs11568500 and rs11568510 in *SLC22A4* were not detected. **Conclusion:** *SLC22A5* variations are involved in the development of IBD, possibly through imbalances in carnitine metabolism. Further biological studies are warranted to determine disease mechanism and to enable reverse phenotyping to determine patients who would benefit from dietary intervention.

4.2 INTRODUCTION

Crohn´s disease (CD) and ulcerative colitis (UC) are the main manifestations of inflammatory bowel disease (IBD)[1]. The disease arises from a complex interplay of environmental, host immune dysregulations and genetic factors[2].

The *IBD5* locus in chromosome *5q31* was first identified to confer CD risk in a Canadian population[3]. This genomic region contains immune related genes: interleukin-4 (IL4), IL13, IL5 and interferon regulatory factor-1 (IRF1), but it also contains two organic cation/carnitine transporters *SLC22A4* and *SLC22A5*. A 250 kb *IBD5* haplotype was associated to CD[4].

It was suggested by Peltekova et al[5] that the non-synonymous SNP rs1050152, located in *SLC22A4* exon 9, and the SNP rs2631367, located in the *SLC22A5* 5´UTR, were true functional polymorphisms determining the CD risk in the *IBD5* locus in a haplotype-independent manner. The associations have been replicated[6,7,8,9,10,11,12,13], but seldom independently from linked SNPs in the *IBD5* haplotype. Moreover, the IBD5 locus including *SLC22A4* and *SLC22A5* has been associated with UC[14,15,16.12], but associations were not replicated in Canadian[6], Belgian[17], and other cohorts of different ancestry[18,19].

A meta-analysis suggests that the SNPs rs12521868 (IGR2096), rs11739135 (IGR2198) and rs17622208 (IGR2230) tag the *IBD5* locus, and are in linkage with *SLC22A4*-rs1050152 and *SLC22A5*-rs2631367 and associated with CD and UC in Caucasian cohorts[20,21].

Taken together there is mounting evidence for the involvement of functional genetic variations in the *SLC22A4* and *SLC22A5* genes, encoding organic (cat)ion transporter

proteins OCTN1 (*SLC22A4*) and OCTN2 (*SLC22A5*). However, it remains undetermined if they act together or independent of each other or independent of nearby variations in immunity-related genes.

SLC22A5 functions as a carnitine transport protein on the apical pole of polarized cells[22]. Carnitine is required for the transport of long-chain fatty acids into the mitochondria and subsequent β-oxidation[23].

SLC22A4 mediates cellular uptake of ergothioneine, which is regarded as a stable antioxidant[24]. Lower blood levels of ergothioneine were found in CD patients compared to healthy controls. *Slc22a4*$^{-/-}$ mice model showed lower tolerance to intestinal oxidative stress[25].

Here we investigated associations of both common and rare variations in the *SLC22A4* and common variations in the *SLC22A5* in a cohort of Caucasian individuals in Manitoba. We included rare variations known to abrogate the proteins function since we hypothesize that they would be observed in the disease cohort if elimination of the gene would have a role in disease development.

4.3 SUBJECTS AND METHODS

4.3.1 Study Population

The study population included 311 IBD patients from the Manitoba Inflammatory Bowel Disease Cohort Study that has been described previously[26]. We included Caucasian age and gender matched CD (n=162), and UC (n= 149) patients as well as healthy controls (n=142). The diagnosis and classification of CD and UC was determined based on radiologic, endoscopic and histological data as established based on the Montreal classification[27]. The phenotypic characteristics of CD and UC patients are shown in Table 2.

4.3.2 Genotyping

All protocols were approved by the University of Manitoba Research Ethics Committee. Genomic DNA was isolated from peripheral blood as described previously[28]. Polymorphisms in *SLC22A4* (rs1050152, rs11568500, rs11568510), in *SLC22A5* (rs17622208) and in *IBD5* locus (rs12521868, rs11739135) were genotyped by PCR-RFLP analysis (General Appendix).

The PCR amplifications were performed in the NEB Taq Polymerase 5X Master Mix (New England BioLabs) following the manufacturers protocol under the following cycling conditions and the primers listed in Table 1: initial denaturation at 95ºC for 30s, followed by 35 cycles of denaturation at 95ºC for 15s, annealing at 50ºC for 15s, extension at 68ºC for 2 min and final extension at 68ºC for 5 min.

The amplicons were digested by allele-specific restriction endonucleases (New England BioLabs) according to manufacturer´s protocols as listed in Table 1. Restriction patterns

were analyzed by gel electrophoresis in a 2% UltraPure agarose gel (Invitrogen) after ethidium bromide staining under UV light (Gel Doc, BIO-RAD). Amplicons of known genotype for every SNP were subcloned using polyAA cloning (TOPO®TA Cloning, Invitrogen) and used as positive and negative controls for further PCR and restriction analysis.

4.3.3 Statistical analyses

SNPs were tested for Hardy-Weinberg equilibrium (HWE). The case-control associations of genotype and allele frequencies of each single nucleotide polymorphisms (SNPs) were tested using binary logistic regression. Odds ratios (OR) were calculated with 95% confidence interval (CI) using 2x2 contingency tables and x^2 test. The analysis was carried out using SPSS 18.0. The linkage disequilibrium (LD) and haplotype analysis in the *IBD5* region was performed with Haploview[29].

4.4 RESULTS

Common SNPs and haplotypes in the *SLC22A5* gene are associated with Crohn's Disease.

Of all interrogated SNPs two singular events in the *SLC22A5* gene were associated with the risk of CD, as well as several haplotypes (Table 3).

The A-allele of SNP rs17622208 was overrepresented in CD patients (OR= 1.4; 95% CI 1.01-1.92; p=0.04), which also resulted in an elevated disease risk for carriers of the genotypes rs17622208-AA/GA (OR= 2.76; 95% CI 1.54-4.95; p=0.001) (Table 3).

Similarly, the C-allele of SNP rs11739135 was overrepresented in the CD patients (OR=1.8; 95% CI 1.28-2.5; p= 0.000) (Figure 2). Significantly, the disease risk for rs11739135-CC homozygotes was strongly elevated (OR=7.84; 95% CI 2.84-21.6; p=0.000) (Figure 2) through the fact that 20.6% of CD patients, but only 3.5% of healthy controls carried that genotype (Table 3).

None of the other two common tag-SNPs in the *IBD5* locus were singularly associated to CD, but the Haplotype TACT was associated with the most elevated risk for CD (OR=1.8; 95% CI 1.07-3.04; p=0.02). The 108 kb IBD5 haplotype block has a high degree of linkage for the CD patients (D´=0.84-0.98) (Figure 1), where the haplotype CGGG conferred the least risk for CD (OR=0.55; 95% CI 0.33-0.93; p=0.02) (Table 3). Other haplotypes were also found to associate, such as haplotype CAGG, which was found in 12.9% in CD compared to 6.3% in controls (p=0.00). Moreover, haplotype TGCT was carried by 4.2% of CD patients compared to 0.3% in controls (p=0.00).

Haplotype TAGT was protective since it was found in 3.2% of CD compared to 9.7% in controls (p=0.00).

CC-homozygosity for common SNP rs11739135 is associated to Ulcerative colitis.

Homozygosity for SNP rs11739135-CC was strongly associated to UC (OR=4.18; 95% CI 1.48-11.78; p=0.007), where the high risk was related to the overrepresentation of the genotype in CD patients (14.8% in CD versus 3.5% in healthy controls) (Table 4).

No other SNP or haplotype had an impact on the risk for UC. The linkage between the common SNPs was significant (D´= 0.85-0.97), but less strong compared to the CD patients (Figure 1).

The rare polymorphism rs11568500-A and rs11568510-G were not present in CD or UC.

The rare functional polymorphisms rs11568500-AA and rs11568510-GG in *SLC22A4* were not detected in CD, UC or the control groups. This Caucasian population carried only the homozygous wild type genotype.

4.5 DISCUSSION

To further characterize the risk alleles in the *SLC22A4* and *SLC22A5* genes, we genotyped common variations in the *SLC22A4*, *SLC22A5* and *C5orf56* genes as well as two rare functional variants in *SLC22A4*. The common alleles rs17622208-A and rs11739135-C in the *SLC22A5* gene associated with CD, and rs11739135-C to UC. None of the common SNPs outside the *SLC22A5* gene associated with IBD, and no individuals in our Manitoban Caucasian cohort carried the rarer *SLC22A4* nonsynonymous alleles rs11568510-G (165Gly) and rs11568500-A (205Ile). Our results confirm previously reported associations for the *SLC22A5* gene, however do not replicate findings for the SNPs in *C5orf56* and *SLC22A4*[20, 21].

***SLC22A4* variations do not contribute to the IBD risk in our cohort.**

The common nonsynonyous SNP rs1050152-T in *SLC22A4* which encodes amino acid 503F was previously reported by Peltekova *et al*[5]. to be present in 53% of CD cases but only 23% of healthy control, indicating a strong disease association, and these findings had been replicated in different cohorts[30,13,31]. However, these associations could not be replicated by others[32,10,17,11], for example Waller et al[15] who found that 26.6% of CD cases and 22.0% of controls carried rs1050152-T. Similarly, we did not find the rs1050152-T associated with IBD in our Manitoban Caucasian cohort where 49% of CD patients, 41% of UC patients, and 43% of controls carried the risk genotype. These findings support the recently formulated hypothesis that the increased rs1050152-T frequency in IBD cases is related to recent positive selection in the IBD5 locus and that other linked disease-causing variants have hitchhiked to relatively high frequency to determine the risk haplotype[33]. Huff et al. postulate that a recombination breakpoint

exists telomeric of *SLC22A4* and that the causative variations are located in the genetic region after that breakpoint, which includes *SLC22A5*. Our data are consistent with this hypothesis, since we did not see disease associations for the *SLC22A4* tag SNP. Moreover, haplotype analysis by Waller et al[15] and Silverberg et al[12] indicated that *SLC22A4* and *SLC22A5* lie in distinct linkage blocks making it feasible to assume that variations in both genes could determine the disease risk independently and that *SLC22A4* is not involved in disease etiology in our cohort.

The assumption that *SLC22A4* is not involved in disease etiology is also supported by the fact that we did not find the two rare *SLC22A4* functional variations rs11568510-G and rs11568500-A, which abrogate transport activity totally or by 50%, respectively[34]. We had chosen to genotype both SNPs due to their proven impact on the proteins function to query the model of "genetic heterogeneity", which postulates that the genetic contribution to complex traits is determined by the abundance of rare genetic variants of high effect on the disease phenotype[35]. We did not find the detrimental genotypes, which also supports the assumption that *SLC22A4* variations do not determine the IBD risk in our cohort.

Common SNPs in *SLC22A5* associated with IBD in the Manitoban Caucasian cohort.

We report that common variations in the *SLC22A5* locus are associated with IBD overall, with strong associations for CD and moderate impact on UC, in our cohort. Our findings differ from reports that both *SLC22A4* and *SLC22A5* SNPs are in linkage and involved in IBD. This might be due to the fact that most studies reported associations for *SLC22A4*/*SLC22A5* haplotypes rs1050152/rs2631372[30,13,31] and rs1050152/rs2631367[5,15],

where the SNPs are located 5' of the *SLC22A5* gene, which is still in a haplotype block with *SLC22A4*. Therefore we assume that the previously reported eQTL-type[5,31,13] associations for *SLC22A5* with IBD are due to SNPs in the *SLC22A4* haplotype block. Considering existing data indicating haplotype breakage between the *SLC22A4* and the *SLC22A5* gene just 5' of *SLC22A5*[15,12] we did choose tag SNPs which are clearly located within *SLC22A5*. This explains why we could achieve distinct and independent associations.

The two SNPs we genotyped in *SLC22A5* and which strongly elevated the ORs for CD and UC are located in intron2 (rs17622208) and distal to the 3'UTR (rs11739135), which makes both unlikely candidates to be the functional causal variation. This suggests that these SNPs tag truly causative variations which most likely affect the proteins' function. This calls for functional biological tests of the known nonsynonymous variations in the gene to determine their potential role in IBD.

Once variations with a functional impact are established, reverse phenotyping could be used to confirm the disease penetrance to evaluate their total contribution in IBD. Based on the fact that the majority of the genes associated to IBD fall in pathways related to immune responses, we hypothesize that the percentage of inflicted individual would be moderate. However, we also postulate that these individuals could be easily identified using simple genotyping and could be treated with effective dietary intervention rather than expensive pharmacological therapies, such as infusions of monoclonal antibodies.

Models of *SLC22A5* involvement in IBD.

A role of carnitine deficiency in intestinal inflammation is supported by the observation that intestinal levels of carnitine are reduced by 90% in Slc22a5$^{-/-}$ knockout mice, which develop spontaneous perforations, micro-abscess, necrotic villi leading to gut atrophy[36]. Neonatal Slc22a5$^{-/-}$ mice showed increased apoptosis of enterocytes and lymphocytes which disturbs the epithelial barrier and initiate the inflammatory process[37].

Oral carnitine supplementation or local carnitine-liposomes administration reduced the inflammation and histological damage in the murine trinitrobenzene sulphonic acid induced colitis[38,39].

Supplementation with propionyl-L-carnitine (1g/day) to UC patients induces improvements in the clinical and endoscopic response[40].

Carnitine is an essential metabolite that regulates the transport of long-chain fatty acids into the mitochondria for the generation of ATP by β-oxidation, particularly important for the enterocytes to maintain normal gut morphology and function. In addition, carnitine may act as an antioxidant inhibiting lipid peroxidation protecting the epithelial barrier against reactive oxygen species[41]. Defects in energy and redox metabolism are linked to impaired pathogen elimination via oxidation burst, and the compromised elimination of pathogens has been proposed to lead to IBD[15].

Involvement of a (cat-)ion imbalance in IBD.

Taken together there is mounting evidence that variations in the organic (cat-)ion transporter genes are risk factors in IBD. Based on the associations with the *SLC22A5* gene, our results suggest that a carnitine imbalance might be involved. However, there is reasonable evidence that *SLC22A4* is involved, whereas substrate(s) and physiological roles are poorly defined. Moreover, SNP rs17309827 in the orphan organic (cat-)ion transporter gene *SLC22A23* has been associated with IBD[42]. Therefore we suggest additional functional studies are important to determine the substrates of these IBD candidate genes in the hope of determining metabolite or nutrient imbalances underlying intestinal inflammation. We conclude that the organic cation transporters of the *SLC22A* family are emerging targets for future individualized genotype-specific dietary intervention in IBD.

4.6 REFERENCES

1. Karp, L. C. & Targan, S. R. in *Inflammatory Bowel Disease* (Directoressor, S. R. T. M., Director, F. S. Md. C. & Advisor, L. C. K. R. P. S.) 3–8 (Wiley-Blackwell, 2010). at <http://onlinelibrary.wiley.com/doi/10.1002/9781444318418.ch2/summary>
2. Fiocchi, C. Inflammatory bowel disease: etiology and pathogenesis. *Gastroenterology* **115,** 182–205 (1998).
3. Rioux, J. D. *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am. J. Hum. Genet.* **66,** 1863–1870 (2000).
4. Rioux, J. D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* **29,** 223–228 (2001).
5. Peltekova, V. D. *et al.* Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat. Genet.* **36,** 471–475 (2004).
6. Newman, B. *et al.* A risk haplotype in the Solute Carrier Family 22A4/22A5 gene cluster influences phenotypic expression of Crohn's disease. *Gastroenterology* **128,** 260–269 (2005).
7. Török, H.-P. *et al.* Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease. *Gut* **54,** 1421–1427 (2005).
8. Noble, C. L. *et al.* The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* **129,** 1854–1864 (2005).

9. Russell, R. K. *et al.* Analysis of the influence of OCTN1/2 variants within the IBD5 locus on disease susceptibility and growth indices in early onset inflammatory bowel disease. *Gut* **55,** 1114–1123 (2006).

10. Martínez, A. *et al.* Association of the organic cation transporter OCTN genes with Crohn's disease in the Spanish population. *Eur. J. Hum. Genet.* **14,** 222–226 (2006).

11. Hradsky, O. *et al.* Two independent genetic factors responsible for the associations of the IBD5 locus with Crohn's disease in the Czech population. *Inflamm. Bowel Dis.* **17,** 1523–1529 (2011).

12. Silverberg, M. S. *et al.* Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. *Eur. J. Hum. Genet.* **15,** 328–335 (2007).

13. Repnik, K. & Potočnik, U. Haplotype in the IBD5 region is associated with refractory Crohn's disease in Slovenian patients and modulates expression of the SLC22A5 gene. *J. Gastroenterol.* **46,** 1081–1091 (2011).

14. Giallourakis, C. *et al.* IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. *Am. J. Hum. Genet.* **73,** 205–211 (2003).

15. Waller, S. *et al.* Evidence for association of OCTN genes and IBD5 with ulcerative colitis. *Gut* **55,** 809–814 (2006).

16. McGovern, D. P. B., Van Heel, D. A., Negoro, K., Ahmad, T. & Jewell, D. P. Further evidence of IBD5/CARD15 (NOD2) epistasis in the susceptibility to ulcerative colitis. *Am. J. Hum. Genet.* **73,** 1465–1466 (2003).

17. Vermeire, S. *et al.* Association of organic cation transporter risk haplotype with perianal penetrating Crohn's disease but not with susceptibility to IBD. *Gastroenterology* **129,** 1845–1853 (2005).

18. Li, M. *et al.* OCTN and CARD15 gene polymorphism in Chinese patients with inflammatory bowel disease. *World J. Gastroenterol.* **14,** 4923–4927 (2008).

19. Tosa, M. *et al.* Lack of association between IBD5 and Crohn's disease in Japanese patients demonstrates population-specific differences in inflammatory bowel disease. *Scand. J. Gastroenterol.* **41,** 48–53 (2006).

20. Wang, J. *et al.* Contribution of the IBD5 locus to inflammatory bowel disease: a meta-analysis. *Hum. Genet.* **129,** 597–609 (2011).

21. Xuan, C. *et al.* Association between OCTN1/2 gene polymorphisms (1672C-T, 207G-C) and susceptibility of Crohn's disease: a meta-analysis. *Int J Colorectal Dis* **27,** 11–19 (2012).

22. Koepsell, H., Lips, K. & Volk, C. Polyspecific organic cation transporters: structure, function, physiological roles, and biopharmaceutical implications. *Pharm. Res.* **24,** 1227–1251 (2007).

23. Tamai, I. *et al.* Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2. *J. Biol. Chem.* **273,** 20378–20382 (1998).

24. Gründemann, D. *et al.* Discovery of the ergothioneine transporter. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 5256–5261 (2005).

25. Kato, Y. *et al.* Gene knockout and metabolome analysis of carnitine/organic cation transporter OCTN1. *Pharm. Res.* **27,** 832–840 (2010).

26. Clara, I. *et al.* The Manitoba IBD Index: evidence for a new and simple indicator of IBD activity. *Am. J. Gastroenterol.* **104,** 1754–1763 (2009).

27. Satsangi, J., Silverberg, M. S., Vermeire, S. & Colombel, J.-F. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* **55,** 749–753 (2006).

28. Cantor, M. J., Nickerson, P. & Bernstein, C. N. The role of cytokine gene polymorphisms in determining disease susceptibility and phenotype in inflammatory bowel disease. *Am. J. Gastroenterol.* **100,** 1134–1142 (2005).

29. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265 (2005).

30. Lin, Z. *et al.* OCTN1 variant L503F is associated with familial and sporadic inflammatory bowel disease. *J Crohns Colitis* **4,** 132–138 (2010).

31. Kabakchiev, B. & Silverberg, M. S. Expression quantitative trait Loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology* **144,** 1488–1496.e3 (2013).

32. Lakner, L. *et al.* IGR2096a_1 T and IGR2198a_1 C alleles on IBD5 locus of chromosome 5q31 region confer risk for Crohn's disease in Hungarian patients. *Int J Colorectal Dis* **24,** 503–507 (2009).

33. Huff, C. D. *et al.* Crohn's disease and genetic hitchhiking at IBD5. *Mol. Biol. Evol.* **29,** 101–111 (2012).

34. Urban, T. J. *et al.* Functional effects of protein sequence polymorphisms in the organic cation/ergothioneine transporter OCTN1 (SLC22A4). *Pharmacogenet. Genomics* **17,** 773–782 (2007).

35. Hegele, R. A. Plasma lipoproteins: genetic influences and clinical implications. *Nat. Rev. Genet.* **10,** 109–121 (2009).

36. Shekhawat, P. S. *et al.* Spontaneous development of intestinal and colonic atrophy and inflammation in the carnitine-deficient jvs (OCTN2(-/-)) mice. *Mol. Genet. Metab.* **92,** 315–324 (2007).

37. Sonne, S., Shekhawat, P. S., Matern, D., Ganapathy, V. & Ignatowicz, L. Carnitine deficiency in OCTN2-/- newborn mice leads to a severe gut and immune phenotype with widespread atrophy, apoptosis and a pro-inflammatory response. *PLoS ONE* **7,** e47729 (2012).

38. Fortin, G. *et al.* L-carnitine, a diet component and organic cation transporter OCTN ligand, displays immunosuppressive properties and abrogates intestinal inflammation. *Clin. Exp. Immunol.* **156,** 161–171 (2009).

39. D'Argenio, G. *et al.* Experimental colitis: decreased Octn2 and Atb0+ expression in rat colonocytes induces carnitine depletion that is reversible by carnitine-loaded liposomes. *FASEB J.* **20,** 2544–2546 (2006).

40. Mikhailova, T. L. *et al.* Randomised clinical trial: the efficacy and safety of propionyl-L-carnitine therapy in patients with ulcerative colitis receiving stable oral treatment. *Aliment. Pharmacol. Ther.* **34,** 1088–1097 (2011).

41. Gülçin, I. Antioxidant and antiradical activities of L-carnitine. *Life Sci.* **78,** 803–811 (2006).

42. 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).

4.7 Tables

Table 1. Primer sequence, restriction enzymes and cutting pattern for RFLP genotyping

Gene	dbSNP	Primer Sequence 5´- 3´	Endonuclease	Cutting pattern (bp)
SLC22A4	rs1050152	Forward: TTGATGTTCTTATGTCCCGG Reverse: TGTGCCCAGCCAACAATATG	MnlI	C: 212+97 bp T: 309 bp
SLC22A4	rs11568500	Forward: ACCTTGGCAACCTACACATC Reverse: TTCAGAGGGTTAGAGGGA	Sau96I	G:168 bp A:85bp
SLC22A4	rs11568510	Forward: TTCCTTGGCAGTGGAATCTG Reverse: GAACAAAAGTGTGTCCAGGT	BspmI	A:312 bp G:203+109 bp
IGR2096	rs12521868	Forward: ATCCTCCATGCTACTGCT Reverse: TGGTGTAGCCAGAGTAGA	DraI	G: 308 bp T:159 + 149 bp
IGR2198	rs11739135	Forward: ACTGGCTCTTTACCTGGGAA Reverse: AACTAGTCCCAACGAGATGA	SfaNI	G: 369 bp C:245 + 124 bp
SLC22A5 IGR2230	rs17622208	Forward: AGGTCTATTCCCAGGGAA Reverse: ACTCAGAAGCTGTCCATC	DdeI	G: 164 + 119 bp A:283 bp

Table 2. Phenotypic characteristics of the Caucasian IBD cohort

	Crohn´s Disease cohort (n=154)	**Ulcerative Colitis cohort (n= 143)**
Gender		
Female	91 (59.1%)	87 (60.8%)
Male	63 (40.9%)	56 (39.2%)
Age at diagnosis		
A1(<16 years)	14 (9.1%)	12 (8.4%)
A2 (16-40 years)	101 (65.6%)	78 (54.5%)
A3 (>40 years)	39 (25.3%)	53 (37.1%)
Location		
L1 (Ileal)	69 (44.8%)	-
L2 (Colonic)	33 (21.4%)	-
L3 (Ileocolonic)	51 (33.1%)	-
L4 (isolated upper disease)	1 (0.6%)	-
E1 (UP limited to rectum)	-	10 (7%)
E2 (Left sided, distal)	-	66 (46.2%)
E3 (extensive, pancolitis)	-	67 (46.9%)
Behaviour		
B1(Inflammatory)	66 (42.9%)	-
B2 (Stricturing)	51 (33.1%)	-
B3 (Penetrating/fistulizing)	37 (24%)	-

Table 3. Genotype and allele frequencies in Crohn´s disease and control subjects

	Crohn´s disease (*n*= 160)	**Controls (*n*=142)**	**OR (95% CI)**	***p***
SLC22A4 rs11568510				
Exon 2				
AA	160 (100%)	142 (100%)	ND	
GG	0 (0%)	0 (0%)	ND	
SLC22A4 rs11568500				
Exon 3				
GG	160 (100%)	142 (100%)	ND	
AA	0 (0%)	0 (0%)	ND	
SLC22A4 rs1050152				
Exon 9	*n*= 160	*n*=142		
CC	42 (26.3%)	51 (35.9%)	Ref.	
CT	79 (49.4%)	61 (43%)	1.57 (0.93-2.66)	0.09
TT	39 (24.4%)	30 (21.2%)	1.58 (0.84-2.95)	0.15
CT + TT	118 (73.8%)	91 (64.1%)	1.60 (0.96-2.57)	0.07
C allele	163 (51%)	163 (57%)	0.77 (0.56- 1.06)	0.11
T allele	157 (49%)	121 (43%)	1.3 (0-94-1.8)	0.11
***SLC22A5* rs17622208**				
Intron 2	*n*= 159	*n*=142		
GG	21 (13.2%)	42 (29.6%)	Ref.	
GA	94 (59.1%)	61 (43%)	3.08(1.67-5.70)	0.000
AA	44 (27.7%)	39 (27.5%)	2.26 (1.14-4.44)	0.019
GA + AA	138 (86.8%)	100(70.4%)	2.76 (1.54-4.95)	0.001
G allele	136 (43%)	145 (51%)	0.72 (0.52-0.98)	0.04
A allele	182 (57%)	139 (49%)	1.4 (1.01-1.92)	0.04
***SLC22A5* rs11739135**				
Intergenic near 3'	*n*= 160	*n*=142		
GG	48 (30%)	57 (40.1%)	Ref.	
GC	79 (49.4%)	80 (56.3%)	1.17 (0.72-1.92)	0.53
CC	33 (20.6%)	5 (3.5%)	7.84 (2.84-21.6)	0.000
GC + CC	112 (70%)	85 (59.8%)	1.56 (0.97-2.52)	0.06
G allele	175 (55%)	194 (68%)	0.56 (0.40-0.78)	0.000
C allele	145 (45%)	90 (32%)	1.8 (1.28-2.5)	0.000
***C5orf56* rs12521868**				
Intron 2	*n*= 159	*n*=142		
GG	43 (27%)	53 (37.3%)	Ref.	
GT	83 (52.2%)	62 (43.7%)	1.65 (0.98-2.77)	0.06
TT	33 (20.8%)	27 (19%)	1.51 (0.78-2.88)	0.22
GT + TT	116 (72.9%)	89 (62.7%)	1.61 (0.98-2.62)	0.06
G allele	169 (53%)	168 (59%)	0.78 (0.57-1.08)	0.13
T allele	149 (47%)	116 (41%)	1.28 (0.92-1.76)	0.13
IBD5 Haplotype[a]				
CGGG	59 (36.4%)	71 (50%)	0.55 (0.33-0.93)	0.02

TACT	63 (38.8%)	42 (29.8%)	1.8 (1.07-304)	0.02
CAGG	21 (12.9%)	9 (6.3%)	-	0.00
TGCT	7 (4.2%)	1 (0.3%)	-	0.00
TAGT	6 (3.2%)	14 (9.7%)	-	0.00

[a]The haplotypes were formed by the SNPs rs1050152, rs17622208, rs11739135, rs12521868.

Table 4. Genotype and allele frequencies in ulcerative colitis and control subjects

	Ulcerative colitis (*n*= 149)	**Controls (*n*=142)**	**OR (95% CI)**	***p***
***SLC22A4* rs11568510**				
Exon 2				
AA	149 (100%)	142 (100%)	ND	
GG	0 (0%)	0 (0%)	ND	
***SLC22A4* rs11568500**				
Exon 3				
GG	149 (100%)	142 (100%)	ND	
AA	0 (0%)	0 (0%)	ND	
***SLC22A4* rs1050152**				
Exon 9				
CC	54 (36.2%)	51 (35.9%)	Ref.	
CT	69 (46.3%)	61 (43%)	1.07 (0.64-1.79)	0.80
TT	26 (17.4%)	30 (21.2%)	0.82 (0.43-1.57)	0.55
CT + TT	95 (63.8%)	91 (64.1%)	0.99 (0.61-1.59)	0.95
C allele	177 (59%)	163 (57%)	1.09 (0.78-1.51)	0.62
T allele	121 (41%)	121 (43%)	0.92 (0.66-1.28)	0.62
***SLC22A5* rs17622208**				
Intron 2				
GG	44 (29.5%)	42 (29.6%)	Ref.	
GA	76 (51%)	61 (43%)	1.19 (0.69-2.04)	0.53
AA	29 (19.5%)	39 (27.5%)	0.71 (0.37-1.35)	0.29
GA + AA	105 (70.5%)	100(70.4%)	1.00 (0.61-1.66)	0.99
G allele	164 (55%)	145 (51%)	1.17 (0.85-1.62)	0.33
A allele	134 (45%)	139 (49%)	0.85 (0.61-1.18)	0.33
***SLC22A5* rs11739135**				
Intergenic near 3'				
GG	60 (40.3%)	57 (40.1%)	Ref.	
GC	67 (45%)	80 (56.3%)	0.79 (0.49-1.29)	0.36
CC	22 (14.8%)	5 (3.5%)	4.18(1.48-11.78)	0.007
GC + CC	89 (59.7%)	85 (59.8%)	0.99 (0.62-1.59)	0.98
G allele	187 (63%)	194 (68%)	0.78 (0.55-1.10)	0.15
C allele	111 (37%)	90 (32%)	1.28 (0.91-1.80)	0.15
***C5orf56* rs12521868**				
Intron 2				
GG	58 (38.9%)	53 (37.3%)	Ref.	
GT	67 (45%)	62 (43.7%)	0.99 (0.59-1.64)	0.96
TT	24 (16.1%)	27 (19%)	0.81 (0.42-1.58)	0.54
GT + TT	91 (61.1%)	89 (62.7%)	0.93 (0.58-1.50)	0.78
G allele	183 (61%)	168 (59%)	1.10 (0.78-1.53)	0.58
T allele	115 (39%)	116 (41%)	0.91 (0.65-1.27)	0.58

***IBD5* Haplotype**[a]				
CGGG	74 (49.4%)	72 (50.2%)	0.96 (0.57-1.63)	0.88
TACT	46 (30.9%)	43 (29.7%)	1.04 (0.61-1.76)	0.88

[a]The haplotypes were formed by the SNPs rs1050152, rs17622208, rs11739135, rs12521868. All SNPs conferred to the Hardy-Weinberg equilibrium.

4.8 Figures

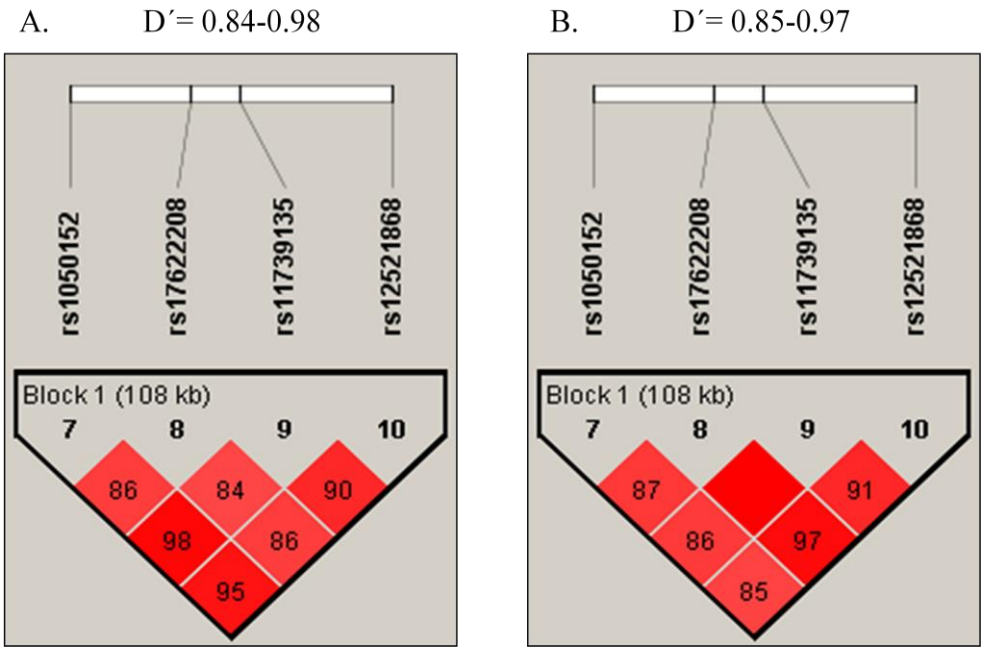


Figure 1. Linkage Disequilibrium of the four genotyped SNPs: the figure shows the coefficient of the linkage disequilibrium (D´) for the studied SNPs in (A) Crohn´s disease and (B) Ulcerative colitis.

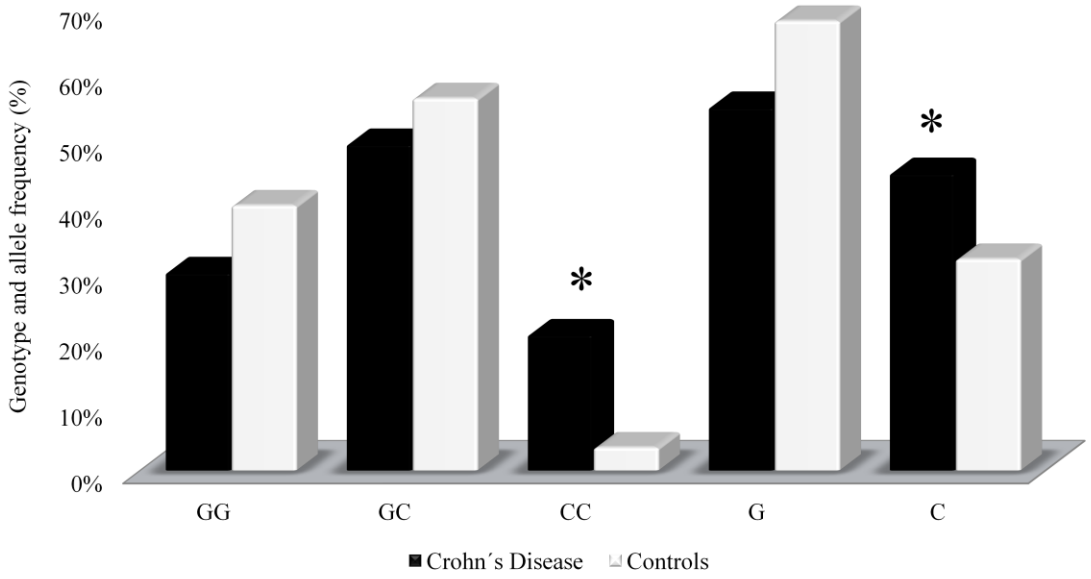


Figure 2. rs11739135 genotype and allele frequencies in Crohn´s disease compared to healthy controls.

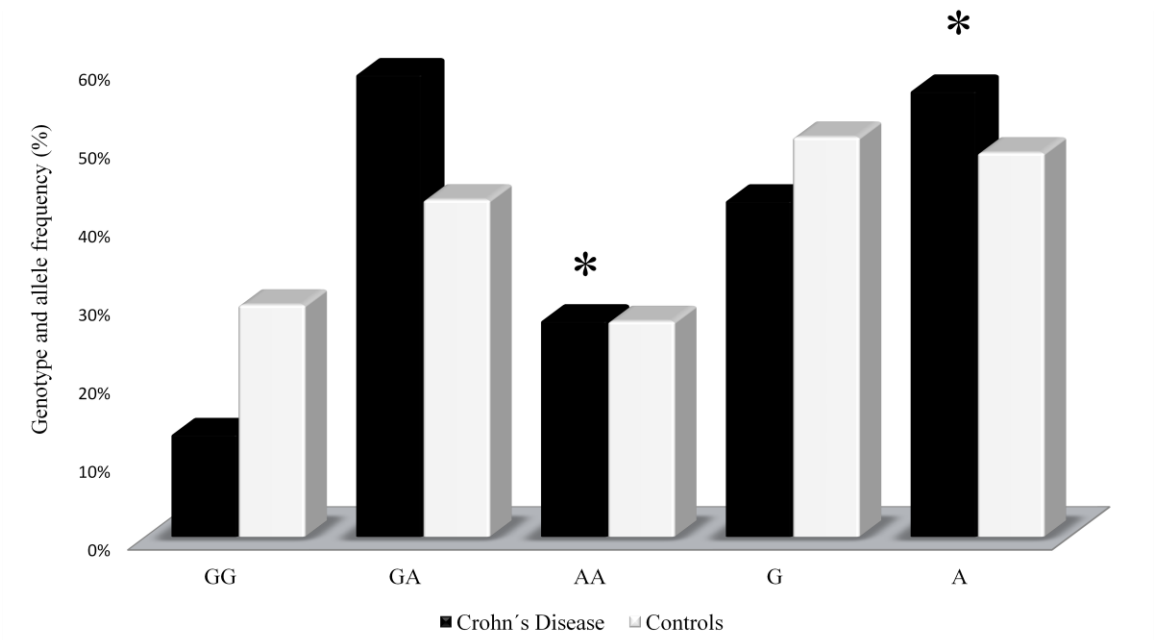


Figure 3. rs17622208 genotype and allele frequencies in Crohn´s disease compared to healthy controls.

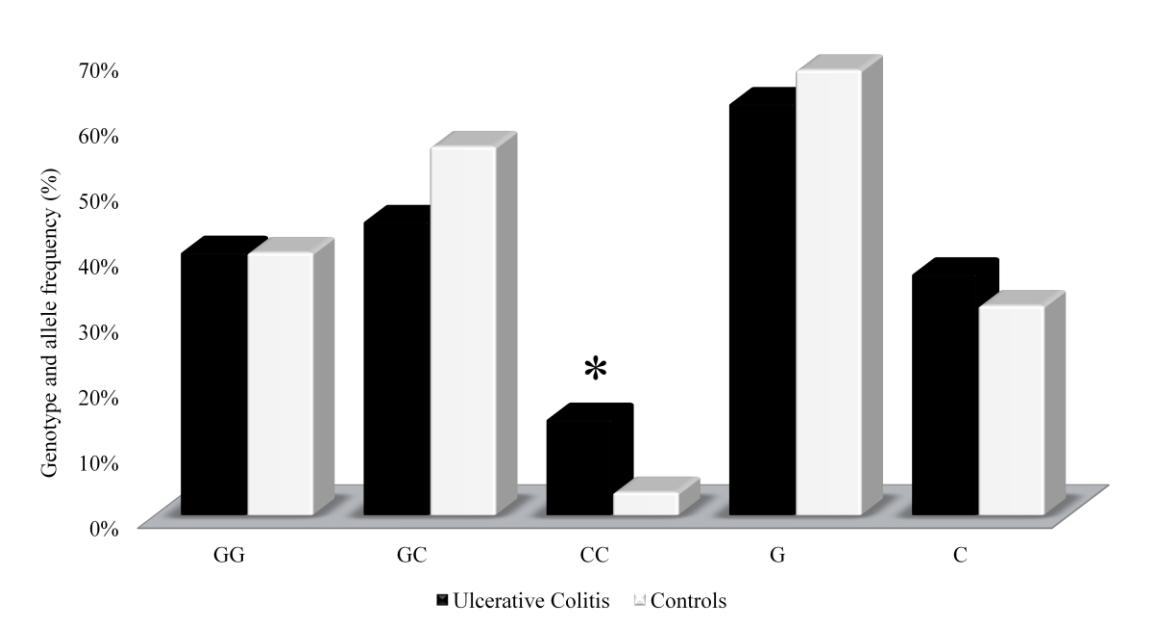


Figure 4. rs11739135 genotype and allele frequencies in ulcerative colitis compared to healthy controls.

CHAPTER 5. Manuscript 2: Single Nucleotide Polymorphisms in *SLC22A23* are associated to Ulcerative Colitis in a Canadian Caucasian Cohort

Alejandra Serrano León[1, 2], Natalia Yurkova[2], Charles N. Bernstein[3,4], Hani El-Gabalawy[4], Peter Eck[1, 2]

[1]Human Nutritional Sciences, [2] The Richardson Centre for Functional Foods and Nutraceuticals, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, [3] University of Manitoba IBD Clinical and Research Centre, [4]Department of Internal Medicine, University of Manitoba, Winnipeg, Canada, R3E 3P4.

5.1 ABSTRACT

Background: *SLC22A23* is an orphan gene in a family of organic membrane transporters. The single nucleotide polymorphism rs17309827-T in *SLC22A23* was recently associated to intestinal inflammation in a genome wide association study. **Objective:** To evaluate if further variations in *SLC22A23* are associated to intestinal inflammation in an inflammatory bowel disease cohort. **Design:** Selected genetic variations were genotyped by TaqMan or PCR-RFLP analysis in 160 Crohn's Disease, 149 Ulcerative Colitis, and 142 healthy controls. Statistical analysis was carried out using SPSS 18.0. Hardy-Weinberg equilibrium and genetic associations were tested using the x^2 test. Odds ratios were determined by 2x2 tables. The linkage disequilibrium (LD) and haplotype analysis was performed in Haploview. **Results:** Homozygotes for the genotypes rs4959235-TT and rs950318-GG were exclusively found in individuals with UC. The presence of genotypes rs4959235-TT and rs950318-GG were more common in UC with Odds Ratios of 11.42 (95% confidence interval (CI) 0.63-266) and 16.71 (0.95-294) respectively. Homozygotes for both genotypes were exclusively found in Crohn's Disease patients, however due to lower numbers statistical significance was not reached. **Conclusion:** For the first time we report strong associations of the two genotypes rs4959235-TT and rs950318-GG to UC. In addition, we hypothesize that these variations could be potential determinants of the overall risk of intestinal inflammation.

5.2 INTRODUCTION

Inflammatory Bowel Disease (IBD) is characterized by chronic inflammation of the gastrointestinal tract, with the principal forms Crohn´s disease (CD) and ulcerative colitis (UC). Mostly unknown environmental and genetic factors contribute to the immune dysregulation, and determine the development, maintenance and severity of the disease[1, 2].

Recent advances in Genome-Wide Association Studies (GWAS) have contributed to the understanding of the genetic factors in intestinal inflammation. Although most genes associated to IBD are in immunoregulatory pathways[3,4], the two genes *SLC22A4* and *SLC22A5*, encoding organic cation transporters, have repeatedly been associated to IBD[5,6,7,8,9,10,11,12,13,14,15,16,17,18]. Recently the SNP rs17309827 located in intron one of the *SLC22A23* gene has been associated to IBD[3,19]. The novel association of *SLC22A23* reinforces the notion that imbalances of organic cations might modulate IBD risk, however, since only one GWAS positively associated the gene independent confirmation is warranted.

SLC22A23 is an orphan gene in the SLC22 family of organic cation and anion transporters, which genomic organization and variations have not been comprehensively described[20] (http://www.ncbi.nlm.nih.gov/gene/63027). Organic cation transporter proteins are involved in the absorption, distribution and excretion of therapeutic drugs, xenobiotics, and food compounds[21]. Functional variations in *SLC22A23* could therefore impact the pharmacokinetic and bioactivity of unknown organic ions which might modulate susceptibility and severity of IBD or other related common and complex diseases.

To date SNPs in *SLC22A23* gene variation have been associated to IBD (rs17309827)[3], in clearance of antipsychotic drugs (rs4959235)[22], and endometriosis-related infertility (specific haplotype)[23]. *SLC22A23* SNP rs17136561 was associated with developing asthma in individuals with impaired allergic status, however, did not reach overall genome wide significance (OR = 1.64; 95% confidence interval 1.33 to 2.02, $P = 2.3\times10^{-6}$)[24]. Moreover, *SLC22A23* was one of six genes which expression levels could be used to predicted the recurrence of triple negative breast cancer in a cohort of Taiwanese woman[25].

The positive association of *SLC22A23* with IBD through GWAS, however, was not replicated in a Caucasian cohort of moderate size[26], highlighting the need for additional replication studies. In this paper the data for gene specific replications in a Caucasian cohort are presented, and positive associations determined.

5.3 SUBJECTS AND METHODS

5.3.1 Study Population

The study population included IBD patients from the Manitoba Inflammatory Bowel Disease Cohort as described previously[27]. A total of 311 age and gender matched Caucasian IBD patients (162 CD, 149 UC) and 142 healthy controls drawn from the general population who did not have personal or first degree relatives with any chronic immune disease were included in the study. CD and UC status was determined based on radiologic, endoscopic and histologic data and phenotype was assigned according to the Montreal classification [28], and are shown in Table 1.

5.3.2 Genotyping

Genomic DNA was isolated from blood as described previously[29]. Polymorphism rs17309827 was genotyped by PCR-RFLP (General Appendix) analysis and rs9503518, rs6923667, rs4959235 were genotyped by TaqMan® SNP genotyping assays (Applied Biosystems®, Foster City, CA).

To determine rs17309827 genotypes, a 219 nt were amplified between the sense primer GGAACGTACAATTCTGCA and the antisense primer GCATGTGAGCGTTTGATG using Taq Polymerase (Taq 5X Master Mix, New England BioLabs) following the manufacturers protocol and the following cycling conditions: initial denaturation at 95ºC for 30s, followed by 35 cycles of denaturation at 95ºC for 15s, annealing at 50ºC for 15s, extension at 68ºC for 2 min and final extension at 68ºC for 5 min. Amplicons were digested with NlaIII (New England BioLabs) using the manufacturers recommendations, where the enzyme cuts the G variant in fragments of 171 nt and 48 nt, while the T variant does not possess a NlaIII restriction site. The restriction patterns of individual samples were determined after gel electrophoresis at a 2% UltraPure agarose gel (Invitrogen) via

ethidium bromide staining under UV light (Gel Doc, BIO-RAD). The efficiency of the amplification and restriction digest was determined by incorporating positive and negative controls amplified from previously subcloned amplicon of known genotype in each set of RFLP analysis.

In addition, rs9503518, rs6923667, rs4959235 were genotype by TaqMan® SNP genotyping assays catalogue number 4351379, 4351379, 4351379, respectively. (Applied Biosystems®, Foster City, CA), using the TaqMan GTXpress Master Mix and the assays as recommended by the manufacturer.

5.3.3 Statistical analyses

Polymorphisms were tested for Hardy-Weinberg equilibrium (HWE). The case-control associations of genotype and allele frequencies of each single nucleotide polymorphisms (SNPs) were tested using binary logistic regression. Odds ratios (OR) were calculated with 95% confidence interval (CI) using 2x2 contingency tables and x^2 test. The analysis was carried out using SPSS 18.0. The linkage disequilibrium (LD) and haplotype blocks was performed with Haploview[30] using the default method[31].

5.3.4 Database analyses

The predicted functional impact of individual SNPs were assessed via the SNP Function Prediction tool FuncPred[32], which includes assessment of functional implications of non-synonymous SNPs, splicing regulation, stop codon changes, PolyPhen, SNPs3D, transcription factor binding sites, microRNA binding sites, regulatory potential, and conservation.

The Hapmap[33]data were assessed via the Ensembl Genome browser[34] and process using Microsoft Excel. All analysis was conducted between July 10th and 22nd 2012.

5.4 RESULTS

Genotypes distributions conform to the Hardy-Weinberg equilibrium for all groups, frequencies are listed in tables 2 and 3. All frequencies observed are in the range reported for Caucasian populations in the HapMap[33] and SNP 500[35], and three linkage blocks were tagged in the 159 kb locus we genotyped (Figure 1).

SNP rs6923667 shows moderate linkage (D´=85 and D´=73) with rs4959235 in CD and in UC, respectively (Figure 1). Three haplotypes are equally distributed in CD (ACCT 27%, ACCG 23.7%, and ATCT 23.6%) and UC (ACCT 26.8%, ATCT 22.9%, and ACCG 22.5%). No specific haplotype was associated with the risk of CD or UC.

The two genotypes rs4959235-TT and rs9503518-GG were exclusively found in UC patients and therefore strongly associated with a markedly elevated risk for the disease (Tables 2, 3). The genotype rs4959235-TT was found in 4% of UC patients but not in healthy individuals and therefore increased the OR to 11.42 (95% CI 0.63-266, p=0.03; Figure 2). The T allele was present in 10.06% of UC patients and 8.09% of healthy controls, but an over transmission in the disease cohort could not be statistically verified.

Genotype rs950318-GG was present in 5.4% of UC patients and none in controls, equating to an OR to 16.71 (95% CI 0.95-294, p=0.00). The G allele was present in 14.8% of UC patients versus 8.4% in healthy controls, indicating a significant over transmission (OR=1.88, 95% CI 1.11-3.18, p=0.01) (Figure 3).

Similar to UC, rs4959235-TT and rs9503518-GG could only be identified in individuals with CD. Of the CD patients 1.9% carried rs4959235-TT and 0.6% carried rs950318-GG (Figure 4), when the allelic frequencies for both SNPs did not differ significantly. In

addition, rs6923667-CT was significantly associated with CD, since 56.8% of patients carried that genotype versus 43% of the controls, elevating the OR to 1.69 (95% CI 1.02-2.77, p=0.04). However, neither frequencies for both homozygotes and alleles differed between the groups.

None of the SNPs were associated to any phenotypic characteristics either in CD or UC (Data not shown).

5.5 DISCUSSION

We demonstrate for the first time that genotypes rs4959235-TT and rs950318-GG in the SLC22A23 gene were strongly associated with UC, in a Caucasian cohort.

For the first time genotypes rs4959235-TT and rs950318-GG in the *SLC22A23* gene were strongly associated to UC, and potentially to IBD in a Caucasian cohort. Overall, both genotypes were exclusively present in IBD patients and not observed in controls, and a statistically significant association was proven for UC, but due to very low frequencies not for CD (Tables 2, 3). Since both genotypes are only present in disease inflicted individuals, the calculated disease risks, as ORs (Tables 2,3) are severely elevated for carriers of these genotypes, resembling values seen for monogenic disorders[36]. Owing to the strongly elevated OR one might argue that these genotypes have the potential to be valid disease predictors or might even functionally contribute to the development of IBD. However, additional independent association studies will be needed to validate these genotypes as disease predictors.

SLC22A23 is a poorly characterized orphan gene and future functional studies are called upon to evaluate potential disease mechanisms. SNP rs4959235 is non-coding in intron four and rs9503518 is coding-synonymous in exon 10. Current prediction tools do not indicate a functional impact of the base changes on transcription factor binding sites, splicing signals, or micro RNA binding sites[32]. However, the current knowledge of *SLC22A23* biology is rudimentary. The gene is in the organic (cat)ion transporter family but the substrate(s) remain undetermined and no regulatory elements are characterized[37].

Since current prediction tools do not indicate any functional impact for SNPs rs4959235 and rs9503518, they could simply be indicators for linked SNPs negatively impacting the genes functions. Currently eight non-synonymous variations and one frame-shift mutation, which are predicted to detrimentally effect the proteins functions, are identified within the linkage region (Table 5)[38]. Further genetic and functional investigations are needed to determine the roles of individual variations in disease development.

Population genetics of rs4959235 indicates a presence of the T-allele in populations with heavier IBD burden.

The IBD risk allele rs4959235-T is not found in Asian and Sub-Saharan populations (HapMap), where IBD incidences are lowest[39]. The rs4959235-T allele reaches moderate allelic frequencies in European cohorts (Table 4), which are known to have a higher IBD burden[40]. Significantly, in four out of five European subpopulations the genotype rs4959235-TT is not documented, despite the fact that 15%-19% of the population is rs4959235-T/C heterozygotes (Table 4). In contrast, rs4959235-TT homozygotes are reported in cohorts of American ancestry, where C/T-heterozygocity is similar to Europeans (Table 4), and this fact might be relevant for IBD development. Unfortunately no information on the associations of rs4959235-TT in American ancestry cohorts is available, but it could be hypothesized that the presence of the rs4959235-TT genotype in these cohorts might indicate a less severe disease impact in TT-carriers compared to European ancestry. This might reflect specific gene-gene interactions defining the IBD risk in the European background, or the fact that rs4959235-T tags nearby variations which negatively impact *SLC22A23* function in a recessive manner in Europeans. This

might at least partially explain the fact that IBD incidence rates are highest in individuals of European/Caucasian ancestry.

Population genetics of rs9503518 are different from rs4959235, which indicates that two distinct linkage blocks in SLC22A23 contribute to IBD susceptibility.

The rs950318-G risk allele is present in all HapMap cohorts regardless of ancestry, and GG-homozygosity ranges from 0.3%-13%. This is in contrast to the findings for the rs4959235-T allele, and validates our observation that both disease associated genotypes are not in genetic linkage (Figure 1). However, as seen for rs4959235-TT carriers, the rs9503518-GG homozygotes were exclusively found in the IBD patients (Tables 2, 3), indicating similar mechanisms of disease genetics, as discussed above.

Significantly, the facts that both disease-associated genotypes tag distinct linkage blocks for the *SLC22A23* gene and are of relatively low frequencies, seem to provide evidence that individual rarer variations of high impact define the disease risk. This supports a genetic "heterogeneity" model, which proposes that genetic contribution to complex traits is comprised of numerous rare genetic variants of high effect on phenotypes, which often are found in the same gene or metabolic pathway [36]. Findings from recent next generation sequencing projects seem to support this model, as the NHLBI-ESP project, the largest exome sequencing project to date, determined 840,000 previously unknown rare variants, of which 164,688 were potentially harmful to a gene and its product [41]. This would mean that an excess of rare genetic variants exists in the human genome and the likelihood of them contributing to a disease process or phenotype is high. The presented data support

the prediction that rarer variations have a higher potential functional impact which result in elevated disease susceptibility in the *SLC22A23* gene [42].

The previously reported disease association of rs17309827-T could not be replicated in the present study.

Previously, the common allele rs17309827-T was associated to CD/UC in a GWAS meta-analysis[3], a finding that was not replicated in this study of a Canadian Caucasian cohort, as it was not replicated in a Caucasian Swedish cohort [26,43,44] and a British UC cohort [45]. Moreover, rs17309827-T has also been associated within a risk stratification model with structuring phenotypes of CD, suggesting the involvement of *SLC22A23* in more severe disease[46]. In our analysis we could not identify any associations to specific disease sub-phenotypes, highlighting the complexity of the disease development in the *SLC22A23* locus, where both common and rarer variants have now been associated with IBD.

IBD associations in SLC22A23 add further evidence that genetic variation in the SLC22A4* and *SLC22A5 genes might be causative in disease development.

Two other genes in the organic (cat)ion transporter family, *SLC22A4* and *SLC22A5* have consistently been associated with IBD[7,9,14,12,11,13,18]. They mediate carnitine and ergothioneine cellular uptake, however, they are in high linkage with the immune-regulatory genes IRF1 and IL5 in the 250kb IBD 5 locus on chromosome 5, and it is

currently debated if variations in *SLC22A4* and *SLC22A5* are "hitchhiking" the true causative variants in IRF1 and IL5 [47]. The newly established associations of *SLC22A23* to IBD support the theory that imbalances in organic (cat)ions, either of endogenous or dietary origin, contribute to IBD susceptibility and that disregulations of *SLC22A4* or *SLC22A5* are disease causative. The disease mechanisms are currently not determined, but it can be assumed that the organic ion imbalances would be caused by functional variations in the *SLC22A* genes.

SLC22A23 in organic ions transport and association to other complex diseases.

Further evidence for the involvement of *SLC22A23* in organic ions transport is established by the association of rs4959235 quetiapine-induced QT prolongation, which is a marker of cardiac arrhythmias in schizophrenia[22]. Quetiapine is an ion compound and the polymorphism in *SLC22A23* gene might be implicated in its absorption and clearance [22]. In addition, the 3´UTR SNPs rs3813486, rs1127473 and rs3211066 form a haplotype, which is associated with infertility and a severe phenotype of endometriosis[23].

The SNP rs9503518 which was associated to UC, could be linked to these 3´UTR haplotypes, suggesting that two very different disease phenotypes could be caused by organic cation imbalances, which modulate immune-pathways[23].

Overall, there is mounting evidence that *SLC22A23*, *SLC22A4* and *SLC22A5* are causatively involved in the susceptibility of IBD, however disease mechanisms remain poorly defined. Moreover, *SLC22A23* is still an orphan transporter gene, calling for functional studies to elucidate its biology. The future identification of the *SLC22A23* substrate(s) will elucidate disease mechanism, with the long term goal to develop

genotype specific dietary interventions. If further validated, genetic variations in this gene could be used as predictive biomarkers for the risk of IBD and to determine dietary or pharmacological therapies.

5.6 REFERENCES

1. Weizman, A. V. & Silverberg, M. S. Have genomic discoveries in inflammatory bowel disease translated into clinical progress? *Curr. Gastroenterol. Rep.* **14,** 139–145 (2012).
2. Podolsky, D. K. Inflammatory bowel disease. *N. Engl. J. Med.* **347,** 417–429 (2002).
3. 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).
4. Lees, C. W., Barrett, J. C., Parkes, M. & Satsangi, J. New IBD genetics: common pathways with other diseases. *Gut* **60,** 1739–1753 (2011).
5. Rioux, J. D. *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am. J. Hum. Genet.* **66,** 1863–1870 (2000).
6. Rioux, J. D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* **29,** 223–228 (2001).
7. Peltekova, V. D. *et al.* Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat. Genet.* **36,** 471–475 (2004).
8. Newman, B. *et al.* A risk haplotype in the Solute Carrier Family 22A4/22A5 gene cluster influences phenotypic expression of Crohn's disease. *Gastroenterology* **128,** 260–269 (2005).
9. Török, H.-P. *et al.* Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease. *Gut* **54,** 1421–1427 (2005).

10. Noble, C. L. *et al.* The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* **129,** 1854–1864 (2005).

11. Russell, R. K. *et al.* Analysis of the influence of OCTN1/2 variants within the IBD5 locus on disease susceptibility and growth indices in early onset inflammatory bowel disease. *Gut* **55,** 1114–1123 (2006).

12. Martínez, A. *et al.* Association of the organic cation transporter OCTN genes with Crohn's disease in the Spanish population. *Eur. J. Hum. Genet. Ejhg* **14,** 222–226 (2006).

13. Hradsky, O. *et al.* Two independent genetic factors responsible for the associations of the IBD5 locus with Crohn's disease in the Czech population. *Inflamm. Bowel Dis.* **17,** 1523–1529 (2011).

14. Silverberg, M. S. *et al.* Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. *Eur. J. Hum. Genet. Ejhg* **15,** 328–335 (2007).

15. Lakner, L. *et al.* IGR2096a_1 T and IGR2198a_1 C alleles on IBD5 locus of chromosome 5q31 region confer risk for Crohn's disease in Hungarian patients. *Int. J. Colorectal Dis.* **24,** 503–507 (2009).

16. Repnik, K. & Potočnik, U. Haplotype in the IBD5 region is associated with refractory Crohn's disease in Slovenian patients and modulates expression of the SLC22A5 gene. *J. Gastroenterol.* **46,** 1081–1091 (2011).

17. Wang, J. *et al.* Contribution of the IBD5 locus to inflammatory bowel disease: a meta-analysis. *Hum. Genet.* **129,** 597–609 (2011).

18. Xuan, C. *et al.* Association between OCTN1/2 gene polymorphisms (1672C-T, 207G-C) and susceptibility of Crohn's disease: a meta-analysis. *Int. J. Colorectal Dis.* **27,** 11–19 (2012).

19. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42,** 1118–1125 (2010).

20. Amberger, J., Bocchini, C. & Hamosh, A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). *Hum. Mutat.* **32,** 564–567 (2011).

21. Tamai, I. Pharmacological and pathophysiological roles of carnitine/organic cation transporters (OCTNs: SLC22A4, SLC22A5 and Slc22a21). *Biopharm. Drug Dispos.* **34,** 29–44 (2013).

22. Aberg, K. *et al.* Genome-wide association study of antipsychotic-induced QTc interval prolongation. *Pharmacogenomics J.* **12,** 165–172 (2012).

23. Zhao, Z. Z. *et al.* Evaluation of polymorphisms in predicted target sites for micro RNAs differentially expressed in endometriosis. *Mol. Hum. Reprod.* **17,** 92–103 (2011).

24. Ramasamy, A. *et al.* Genome-wide association studies of asthma in population-based cohorts confirm known and suggested loci and identify an additional association near HLA. *Plos One* **7,** e44008 (2012).

25. Chen, L. H. *et al.* Identification of prognostic genes for recurrent risk prediction in triple negative breast cancer patients in Taiwan. *Plos One* **6,** e28222 (2011).

26. McGovern, D. P. B. *et al.* Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Hum. Mol. Genet.* **19,** 3468–3476 (2010).

27. Clara, I. *et al.* The Manitoba IBD Index: evidence for a new and simple indicator of IBD activity. *Am. J. Gastroenterol.* **104,** 1754–1763 (2009).

28. Satsangi, J., Silverberg, M. S., Vermeire, S. & Colombel, J.-F. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* **55,** 749–753 (2006).

29. Cantor, M. J., Nickerson, P. & Bernstein, C. N. The role of cytokine gene polymorphisms in determining disease susceptibility and phenotype in inflammatory bowel disease. *Am. J. Gastroenterol.* **100,** 1134–1142 (2005).

30. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinforma. Oxf. Engl.* **21,** 263–265 (2005).

31. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296,** 2225–2229 (2002).

32. Xu, Z. & Taylor, J. A. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res.* **37,** W600–605 (2009).

33. International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).

34. Flicek, P. *et al.* Ensembl 2013. *Nucleic Acids Res.* **41,** D48–D55 (2012).

35. Packer, B. R. *et al.* SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res.* **34,** D617–621 (2006).

36. Hegele, R. A. Plasma lipoproteins: genetic influences and clinical implications. *Nat. Rev. Genet.* **10,** 109–121 (2009).

37. Bennett, K. M., Liu, J., Hoelting, C. & Stoll, J. Expression and analysis of two novel rat organic cation transporter homologs, SLC22A17 and SLC22A23. *Mol. Cell. Biochem.* **352,** 143–154 (2011).

38. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).

39. Yang, S. K., Loftus, E. V., Jr & Sandborn, W. J. Epidemiology of inflammatory bowel disease in Asia. *Inflamm. Bowel Dis.* **7,** 260–270 (2001).

40. Molodecky, N. A. *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* **142,** 46–54.e42; quiz e30 (2012).

41. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337,** 64–69 (2012).

42. Lin, W.-Y. *et al.* Haplotype Kernel Association Test as a Powerful Method to Identify Chromosomal Regions Harboring Uncommon Causal Variants. *Genet. Epidemiol.* (2013). doi:10.1002/gepi.21740

43. Törkvist, L. *et al.* Analysis of 39 Crohn's disease risk loci in Swedish inflammatory bowel disease patients. *Inflamm. Bowel Dis.* **16,** 907–909 (2010).

44. McGovern, D. P. B. *et al.* Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat. Genet.* **42,** 332–337 (2010).

45. Anderson, C. A. *et al.* Investigation of Crohn's disease risk loci in ulcerative colitis further defines their molecular relationship. *Gastroenterology* **136,** 523–529.e3 (2009).

46. Henckaerts, L. *et al.* Genetic risk profiling and prediction of disease course in Crohn's disease patients. *Clin. Gastroenterol. Hepatol. Off. Clin. Pr. J. Am. Gastroenterol. Assoc.* **7,** 972–980.e2 (2009).

47. Huff, C. D. *et al.* Crohn's disease and genetic hitchhiking at IBD5. *Mol. Biol. Evol.* **29,** 101–111 (2012).

5.7 Tables

Table 1. Phenotypic characteristics of the Caucasian IBD cohort

	Crohn´s Disease cohort (n=154)	**Ulcerative Colitis cohort (n= 143)**
Gender		
Female	91 (59.1%)	87 (60.8%)
Male	63 (40.9%)	56 (39.2%)
Age at diagnosis		
A1(<16 years)	14 (9.1%)	12 (8.4%)
A2 (16-40 years)	101 (65.6%)	78 (54.5%)
A3 (>40 years)	39 (25.3%)	53 (37.1%)
Location		
L1 (Ileal)	69 (44.8%)	-
L2 (Colonic)	33 (21.4%)	-
L3 (Ileocolonic)	51 (33.1%)	-
L4 (isolated upper disease)	1 (0.6%)	-
E1 (UP limited to rectum)	-	10 (7%)
E2 (Left sided, distal)	-	66 (46.2%)
E3 (extensive, pancolitis)	-	67 (46.9%)
Behaviour		
B1(Inflammatory)	66 (42.9%)	-
B2 (Stricturing)	51 (33.1%)	-
B3 (Penetrating/fistulizing)	37 (24%)	-

Table 2. Association of *SLC22A23* SNPs with Crohn´s Disease

	Crohn´s disease	**Controls**	**OR (95% CI)**	***p***
rs4959235	***n*= 162**	***n*=142**		
CC	141 (87%)	119 (83.8%)	Ref.	
CT	18 (11.1%)	23 (16.2%)	0.66 (0.34-1.28)	0.22
TT	3 (1.9%)	0 (0%)	5.06 (0.25-1.02)	0.24
C allele	300 (92.5%)	261 (91.9%)	1.10 (0.60-1.99)	0.75
T allele	24 (7.4%)	23 (8.09%)	0.91 (0.50-1.64)	0.75
rs6923667	***n*= 162**	***n*=142**		
CC	51 (31.5%)	57 (40.1%)	Ref.	
CT	92 (56.8%)	61 (43%)	1.69 (1.02-2.77)	0.04
TT	19 (11.7%)	24 (16.9%)	0.88 (0.44-1.80)	0.74
C allele	194 (59.8%)	175 (61.6%)	0.92 (0.67-1.28)	0.66
T allele	130 (40.1%)	109 (38.3%)	1.08 (0.78-1.49)	0.66
rs9503518	***n*= 162**	***n*=142**		
AA	139 (85.8%)	118 (83.1%)	Ref.	
AG	22 (13.6%)	24 (16.9%)	0.78 (0.42-1.46)	0.43
GG	1 (0.6%)	0 (0%)	1.69 (0.06-51)	0.75
A allele	300 (92.5%)	260 (91.5%)	1.15 (0.64-2.08)	0.63
G allele	24 (7.4%)	24 (8.4%)	0.87 (0.48-1.56)	0.63
rs17309827	***n*= 160**	***n*=142**		
GG	21 (13.1%)	20 (14.1%)	Ref.	
GT	81 (50.6%)	70 (49.3%)	1.10 (0.55- 2.19)	0.78
TT	58 (36.3%)	52 (36.6%)	1.06 (0.52-2.17)	0.87
G allele	123 (38%)	110 (39%)	0.99 (0.71-1.37)	0.94
T allele	197 (62%)	174 (61%)	1.01 (0.73-1.41)	0.94

Table 3. Association of *SLC22A23* SNPs with Ulcerative Colitis

	Ulcerative Colitis	**Controls**	**OR (95% CI)**	***p***
rs4959235	***n*= 149**	***n*=142**		
CC	125 (83.9%)	119 (83.8%)	Ref.	
CT	18 (12.1%)	23 (16.2%)	0.74 (0.38-1.45)	0.39
TT	6 (4.0%)	0 (0%)	11.42 (0.63-266)	0.03
C allele	268 (89.9%)	261 (91.9%)	0.78 (0.44-1.39)	0.40
T allele	30 (10.06%)	23 (8.09%)	1.27 (0.72-2.24)	0.40
rs6923667	***n*= 149**	***n*=142**		
CC	62 (41.6%)	57 (40.1%)	Ref.	
CT	66 (44.3%)	61 (43%)	0.99 (0.60-1.64)	0.98
TT	21 (14.1%)	24 (16.9%)	0.80 (0.40-1.60)	0.53
C allele	190 (63.7%)	175 (61.6%)	1.09 (0.78-1.53)	0.59
T allele	108 (36.2%)	109 (38.3%)	0.91 (0.65-1.27)	0.59
rs9503518	***n*= 149**	***n*=142**		
AA	113 (75.8%)	118 (83.1%)	Ref.	
AG	28 (18.8%)	24 (16.9%)	1.22 (0.66-2.22)	0.52
GG	8 (5.4%)	0 (0%)	16.71 (0.95-294)	0.00
A allele	254 (85.2%)	260 (91.5%)	0.53 (0.31-0.90)	0.01
G allele	44 (14.8%)	24 (8.4%)	1.88 (1.11-3.18)	0.01
rs17309827	***n*= 149**	***n*=142**		
GG	23 (15.4%)	20 (14.1%)	Ref.	
GT	60 (40.3%)	70 (49.3%)	0.74 (0.37-1.49)	0.40
TT	66 (44.3%)	52 (36.6%)	1.10 (0.55-2.22)	0.78
G allele	106 (36%)	110 (39%)	0.87 (0.62-1.22)	0.43
T allele	192 (64%)	174 (61%)	1.14 (0.82-1.60)	0.43

Table 4. rs4959235 allele and genotype frequencies in selected HapMap cohorts

Ancestry	Population[a]	Allele		Genotype		
		C	T	C|C	C|T	T|T
European	CEU	0.912	0.088	0.824	0.176	
	FIN	0.898	0.102	0.806	0.183	0.011
	GBR	0.927	0.073	0.854	0.146	
	IBS	0.929	0.071	0.857	0.143	
	TSI	0.98	0.02	0.959	0.041	
American	CLM	0.867	0.133	0.75	0.233	0.017
	MXL	0.856	0.144	0.758	0.197	0.045
	PUR	0.918	0.082	0.836	0.164	
Sub Saharan	LWK	1		1		
	YRI	1		1		
African in the US	ASW	0.951	0.049	0.918	0.066	0.016
Asian	JPT	1		1		
	CHS	1		1		
	CHB	1		1		

[a]CEU:Uta residents with Northern and Western European ancestry; FIN: Finish in Finland; GBR: British in England and Scotlan; IBS: Iberian population in Spain; TSI: Toscani in Italy; CLM: Colombian from Medellin; MXL: Mexican ancestry from Los Angeles USA; PUR: Puerto Ricans from Puerto Rico; LWK: Luhya in Webuye, Kenia; YRI: Yoruba in Ibadan, Nigeria; ASW: Americans of African ancestry in SW USA; JPT: Japanese in Tokyo, Japan; CHS: Southern Han, Chinese; CHB: Han Chinese in Beijing, China[33, 34].

Table 5. Non-synonymous and frameshift polymorphisms linked to rs9503518

dbSNP rs No.	DbSNP allele	Protein residue	Amino acid position
rs200011775	A > G	Thr > Ala	579
rs148262614	T > C	Ser > Pro	587
rs201540017	G > A	Asp > Asn	632
rs199592727	T > C	Cys > Arg	638
rs141223516	C > A	Pro > His	644
rs201396552	A > G	Asn > Asp	666
rs150444727	A > G	Thr > Ala	670
rs112065429	A > G	Ser > Gly	671
rs34106449	Frameshift	-	634

5.8 Figures

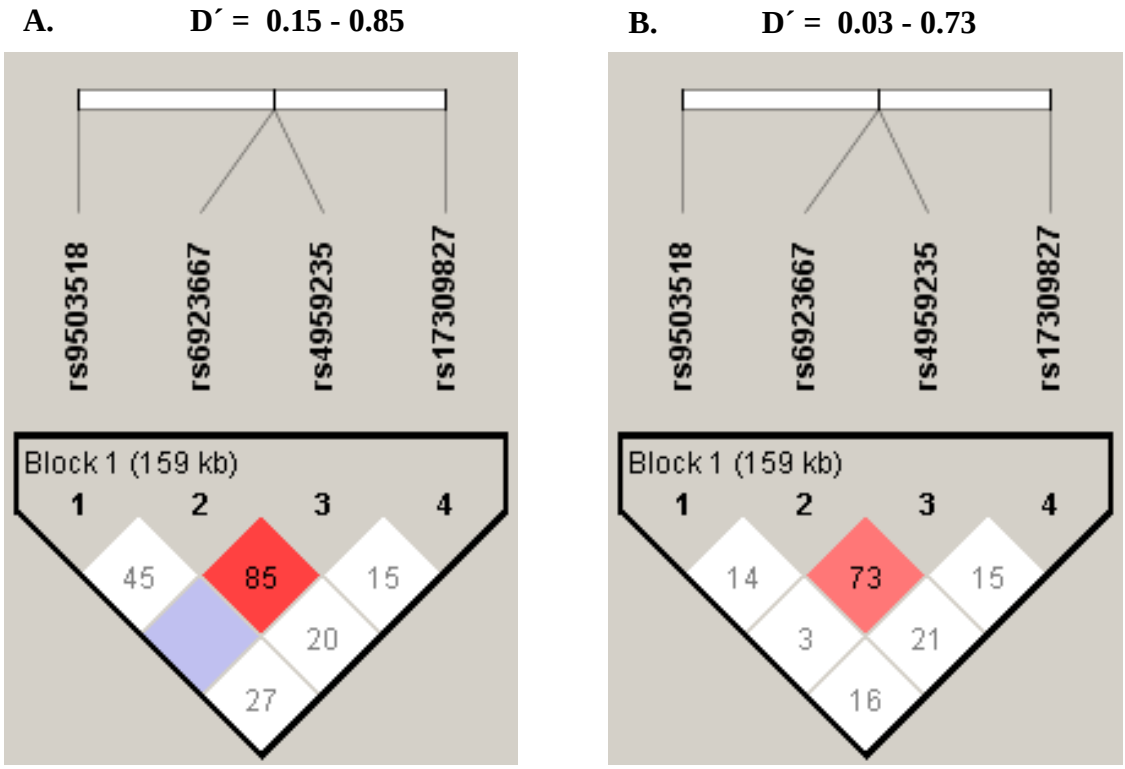
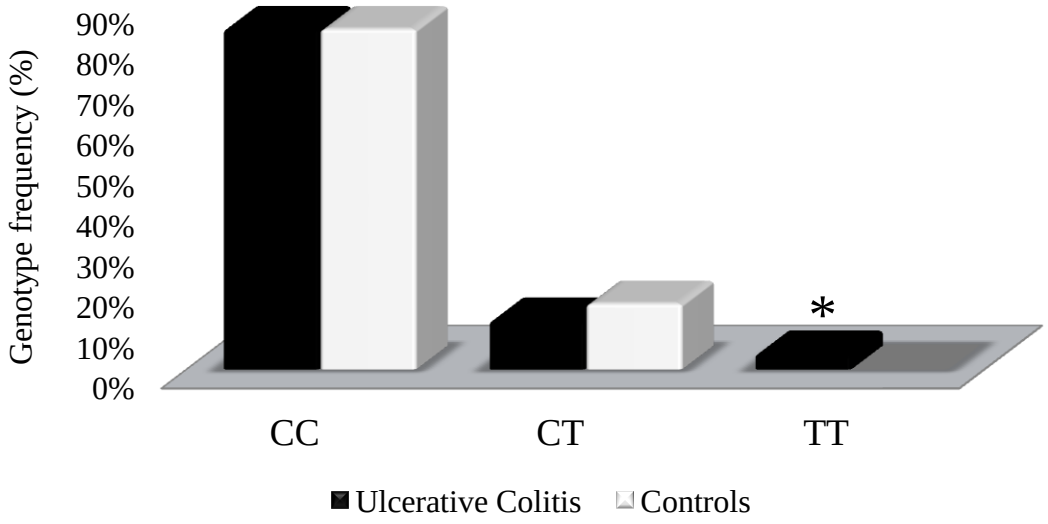
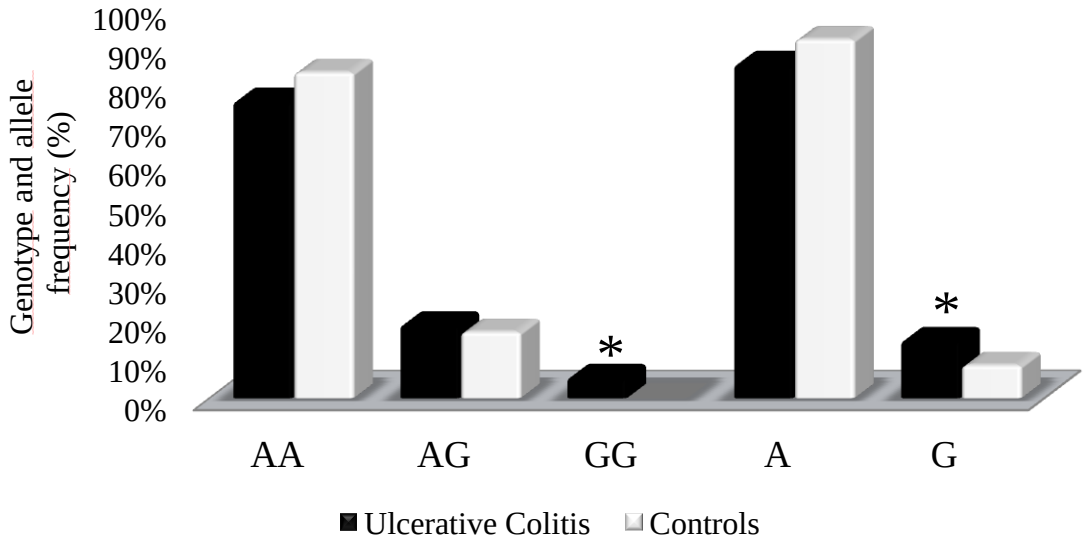


Figure 1. Linkage Disequilibrium of the four genotyped SNPs: the figure shows the coefficient of the linkage disequilibrium (D´) for the studied SNPs in (A) Crohn´s disease and (B) Ulcerative colitis.



* $p = 0.03$

Figure 2. *SLC22A23* rs4959235 genotype frequencies in ulcerative colitis and controls.



* p= 0.00

Figure 3. *SLC22A23* rs9503518 genotype and allele frequencies in ulcerative colitis and controls.

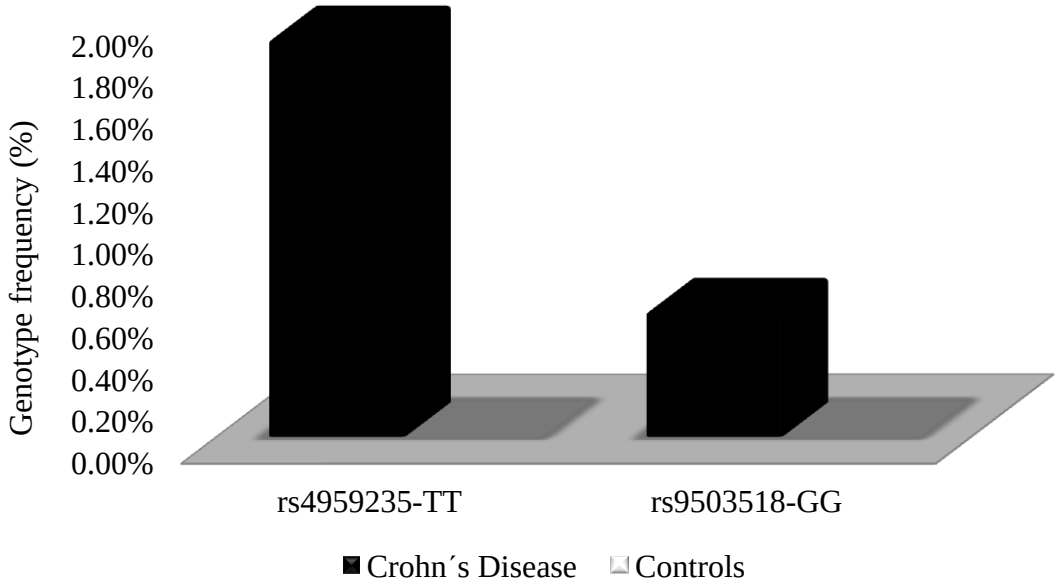


Figure 4. *SLC22A23* rs4959235-TT and rs9503518-GG genotype frequencies in Crohn´s disease and controls.

CHAPTER 6. Manuscript 3: Characterization and expression of human *SLC22A23* gene

Alejandra Serrano León[1, 2], Natalia Yurkova[2], Peter Eck[1, 2]

[1]Human Nutritional Sciences, [2] The Richardson Centre for Functional Foods and Nutraceuticals, University of Manitoba, Winnipeg, Manitoba, R3T 2N2

6.1 ABSTRACT

Background: Organic cation transporters belong to the solute carrier transporter family (SLC22). They are responsible for the absorption of metabolites, food components, drugs or xenobiotics. Recently, the human orphan transporter *SLC22A23* was identified through data base analysis and was classified within the phylogenetic tree of organic cation transporters. Polymorphisms in this gene have been associated to Crohn´s disease. However, the genomic organization and characterization of the *SLC22A23* isoform has not been reported. **Objective:** This study characterizes the structure of the gene and its splice variants. Once identified the major splice variants were subcloned to express the two main isoforms of human *SLC22A23* protein for subsequent cellular imaging and functional studies. **Design:** The genomic structure and transcript variants of *SLC22A23*, were identified via *in silico* database analyses. Transcript variants were subcloned from placenta cDNA and fused to N-terminal fluorescent proteins. Through expression in HEK 293 cells the cellular localization was determined using fluorescent microscopy. **Results:** The the *SLC22A23* gene occupies a 187.59 kb locus and consist of 10 exons. 7 splice variants could be identified, of which two were subcloned. Upon expression in mammalian cells membrane localization was observed. **Conclusion:** The membrane localization makes *SLC22A23* a target gene in intestinal inflammation, where functional variations in ion transporters are associated to disease development. However, it will be essential to determine *SLC22A23s* physiological substrates to allow conclusions on disease mechanisms.

6.2 INTRODUCTION

Organic cation transporters play an important role during the absorption, distribution and excretion of metabolites, food components, drugs or xenobiotics. These transporters belong to the solute carrier family SLC22A [1]. Three types of transporters are included in this family, the organic cation transporters (OCTN1-3), the organic anion transporters (OAT), and carnitine/ organic cation transporters (OCTN1-3)[2].

These transporters share around 50% identities have typically between 500 and 600 amino acids which form 12 α-helical transmembrane domains (TMDs). They have a glycosylated extracellular loop in TMD 1 and 2, and a phosphorylated intracellular loop between the TMD 6 and 7. The N and C-termini are located intracellularly[3].

Recently, the orphan human *SLC22A23* gene was identified through its sequence homology to the rodent *slc22a23*. Its genome structure lead to its classification within the phylogenetic tree of SLC22 transporters[4].

The human *SLC22A23* gene has been implicated in human diseases through genome wide associations with inflammatory bowel disease[5] and clearance of antipsychotic drugs[6]. However, disease relevance remains uncertain, since the gene and its products have not been characterized. Studies of the rat ortholog did not identify a physiological substrate[7].

Here the human gene including its two main transcript splice variants are characterized.

6.3 MATERIALS AND METHODS

6.3.1 Bioinformatics analysis to determine the *SLC22A23* gene structure and expression.

The information was sourced from the NCBI[8] such as PubMed[9], UniGene[10], dbSNP[11] as well as Ensembl[12] databases between January and June 2013. Human genetic sequence for the *SLC22A23* gene were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/ gene/63027)[13] in FASTA files from human RefSeqs 37.3. Expressed sequence tags (ESTs) were sourced from UniGene, NCBI through different BLAST searches and the Ensembl data base (http://uswest.ensembl.org/ ENSG00000137266). Reference transcripts were identified as NM_015482.1 and NM_021945.5.

Sequences were processed and aligned using Sequencher4.8 (Gene Codes) and sequences and alignments were curated through visual inspection.

The *in silico* tissue expression was determined using the UniGene EST profile for cluster Hs. 713588.

6.3.2 Bioinformatics analysis to characterize the *SLC22A23* protein isoforms.

Protein comparisons, alignments and predictions of structures and motifs were performed using the following online software packages and browsers: ProtParm[14], T-COFEE 9.03[15], TM-COFFEE, TMHMM Server v.2.0[16], TOPCONS[17] and Prosite[18].

6.3.3 Sub-Cloning of the human SLC22A23 transcripts

All PCR reactions were performed using Phusion polymerase (New England Biolabs) in the high fidelity buffer. All PCR conditions and primers used are listed in the appendix 1.

Differentially spliced human *SLC22A23* transcripts were amplified from placenta (Human Placenta Marathon-ready cDNA, Clontech) or duodenal (Human duodenum, Clontech) cDNA. Two different transcripts were obtained, which we called isoform A and B. In a second round of PCR they were fused to N-terminal fluorescent proteins, where isoform A was fused to the red fluorescent proteins mCherry and isoform B to the green fluorescent protein tGFP. The terminal ends of the amplicons were tagged with the Gateway® recombination sites AttB1 and AttB2, integrated into the donor vector pDONR2***21*** and transferred into the expression vector pcDNA/V5-DEST. The identities of the inserts were confirmed via sequencing (The Centre for Applied Genomics (TCAG), Toronto).

6.3.4 Cell Culture, Transfection, and visualization of intracellular localization.

HEK293 cells were cultured at 37^{o}C in DMEM/F-12 HyClone media (Thermo Fisher Scientific), 10% FBS on 75mm flasks (Thermo Fisher Scientific, Corning) or 24-well tissue culture plates (Thermo Fisher Scientific, BioLite). Twenty-four hours after seeding and at cell densities between 60%-89% cells were transfected with the plasmid of choice using Effectine transfection reagent following the manufacturer´s protocols (QIAGEN).

Between 24 to 48h post transfection the cells were visualized by fluorescent microscopy on an Axiovert 200 (ZEISS). Images were collected and processed with the ZEN 2012 lite software.

6.4 RESULTS

6.4.1 Location, structure, and transcript variants of the *SLC22A23* gene

The *SLC22A23* gene is located in the chromosome 6p25.2, bases 3,269,196 - 3,457,256 on the reverse strand (Appendix Figure 1A), flanked by the protein assembly chaperon 4 (PSMG4) and the LOC643327 genes. The two reference transcripts NM015482.1 (NP056297.1, isoform A) and NM021945.5 (NP068764.3, isoform B) span a locus of 187.59 kb (Appendix Figure 1B). Both transcripts differ in the utilization of three 5'exons.

NM015482.1 contains 6159 bp and is organized into 10 exons. It utilizes an alternative first exon, which we call exon 1A since it is the most 5' exon in the *SLC22A23* locus (Figure 1A). Exon 1 A is 654 bp long and contains a translation initiation codon (ATG) 444 bp downstream from the transcriptional start point.

NM021945.5 is 5916 bp long and is organized into 11 exons. It utilizes an alternative first exon (exon 1B in figure 1B) and an alternative second exon (exon 2B in figure 1B), before it splices into exon 2A, which is shared by both transcripts. The translation initiation codon (ATG) starts at base 86 in exon 3 (Figure 1B), and all exons thereafter are shared between the reference transcripts. EST evidence (EST Hs#S29618082) indicates a transcriptional start point 30bp upstream of the start currently annotated for NM021945.5, which increases the size of exon 1B to 196 bp and the length of the whole transcript to 5946 bp (Appendix Figure 2).

6.4.2 Additional splice variants in the *SLC22A23* gene.

The alignment of ESTs homologue to the *SLC22A23* locus revealed nine additional splicing events (Figure 2):

1. Exon 2B is skipped.
2. A combined event where the last 151 bp of exon 1A are skipped, exon 2A is skipped, and the first 21 bp of exon 3A are skipped. This splice variant would conform to a GC-TG splice donor-acceptor event (Appendix Figure 3).
3. An alternative 164bp exon 3B could be utilized in intron 3 (Appendix Figure 4).
4. An alternative 86 bp exon 3C could be utilized in intron 3 (Appendix Figure 5).
5. Last 53 bp of exon 4 are skipped.
6. An alternative 53 bp exon 4B could be utilized in intron 4 (Appendix Figure 6).
7. An alternative 296 bp exon 9B could be utilized in intron 9 (Appendix Figure 7).
8. Exon 9 is elongated through intron retention of the last 145 bases of intron 8.
9. 337 bp in exon 10 are deleted.

Two alternative transcription termination sites are utilized in exon 10 (Figure 3).

The analysis also revealed the existence of a TCA insertion in the 3´UTR, corresponding to rs5873875 in dbSNP (Appendix Figure 8).

6.4.3 Tissue expression of *SLC22A23*

SLC22A23 transcripts are found in a wide variety of tissues, with highest expression in bladder, cervix, intestine and lung. The high expression in the intestine seem to be of relevance, since a variation in the gene was recently associated to intestinal inflammation[5]. Lower expression is reported for liver, kidney, heart, brain, nerve, pancreas, eye, blood, esophagus, stomach, trachea, testis muscle, ovary, parathyroid, prostate, skin and bone. Significant expression is mapped to cervical, colon,

gastrointestinal, liver, lung, and prostate tumors, indicating a possible involvement in cancer etiology.

6.4.4 The *SLC22A23* protein isoforms

The two reference protein isoform NP056297.1, also called isoform A, and NP068764.3, also called isoform B, differ in their N-terminal sequence. The longer isoform A is predicted to encode 686 amino acids with a molecular weight of 73748.2 Da, the theoretical pI is 7.98 and the ratio of negatively charged to positively charged amino acid is 50/53 (sequence in the appendix). Isoform B encodes a smaller protein of 405 amino acids with a molecular weight of 44784.2 Da, the theoretical pI is 8.67 and shows a ratio of negatively charged to positively charged amino acid of 28/35[14] (sequence in the appendix). Isoform A has 281 additional amino acids in the N-terminus (alignment in the appendix).

The predicted membrane architecture differs remarkably, with isoform A having a cytosolic and isoform B an extracellular N-terminus. According to TM-COFEE[15] and TOPCONS[17], isoform A is predicted to have 12 TMD and only 9 TMD are predicted in isoform B. The C-terminus of both proteins is cytosolic. However, TMHMM[16] predicts 10 TMD for isoform A and 9 TMD for isoform B (Figure 4).

Based on sequence homology (Prosite[18]) isoform A is identified as a member of the Major Facilitator Superfamily proteins. Moreover, it has a sugar transport site, 3 amidation sites, 17 myristyl sites, 6 glycosylation sites, 8 casein kinase II phosphorylation sites (CK2), 7 protein kinase C phosphorylation sites (PKC) and a nuclear localization signal (GGRAGPGGG).

Isoform B has 2 amidation sites, 7 myristyl sites, 1 glycosylation site, 4 casein kinase II phosphorylation sites (CK2), 6 protein kinase C phosphorylation sites (PKC). It also has sugar transport site but the Major Facilitator Superfamily motif was not identified[18].

6.4.5 Expression of human *SLC22A23* isoform B in HEK293 cells

To evaluate a possible function as a transmembrane transporter protein on the plasmalemma membrane tGFP tagged *SLC22A23* isoform B was transiently expressed in HEK cells. We focussed on isoform B, since isoform A was predicted to exhibit a nuclear localization.

Green fluorescence clearly located to the plasmalemma membrane as shown in Figure 5.

6.5 DISCUSSION

Here we characterize, for the first time, the human *SLC22A23* locus and predict its alternative transcripts. The human gene is still orphan, belonging to the *SLC22A* family of organic ion membrane transporters, but has never been characterized on the genetic or protein level before. Transporters in the *SLC22A* family are thought to be uniporters, symporters and antiporters of organic ions[4]. Our work represents the first step towards the identification of the physiologic substrate, studies that are warranted due to the fact that genetic variations in the *SLC22A23* locus are associated to intestinal inflammation (rs17309827)[5], in clearance of antipsychotic drugs[6] and endometriosis-related infertility[19].

Our cellular expression studies of the major protein isoform B shows a plasmamembrane localization (Figure 5), confirming that the human *SLC22A23* gene encodes a membrane protein and making it feasible to assume that it is involved in organic cation transport. Similar membrane localization is reported for the rat homolog Boct2[7], but transport studies on the rat homolog failed to identify an elevated cellular uptake for 1-methyl-4-phenylpyridinium (MPP^+), tetraethylammonium (TEA), carnitine, choline, histamine, Norepinephrine, Nicotinamide, Urate, and Ascorbate[7]. Additional studies are warranted, to determine the physiologic substrate and to elucidate the biology underlying the disease associations.

Our *in silico* analysis reveals significant alternative splicing within the human *SLC22A23* gene, which might contribute to disease susceptibilities. Splicing is one of the processes occurring in cells to regulate gene expression at the RNA processing level. The splicing process is variable within a gene, thus different sites can be used to produce alternative

splice mRNA isoforms[20]. The alternative isoforms can be translated into proteins with different functions or they may have similar function but being regulated differently according to tissue-specificity or cell differentiation[21]. Alternative splicing also mediate regulatory elements to control translation, mRNA stability or localization. Mutations that result in dysregulation of alternative splicing can cause the expression of isoforms at inappropriate tissues or developmental stages. This process is also being considered to explain genetic disorders which lead to disease conditions[22]. Within the organic cation transporter family genes alternative splicing is linked to detrimental phenotypes, such as the splice isoform OCTN2-VT (*SLC22A5*), where an addition of 72 base pairs to the first exon encodes a protein with additional 24 amino acids, leading to poor N-glycosylation and retention in the endoplasmic reticulum. This alternative SLC22A5 isoform did not mediate carnitine uptake, which is the main substrate for the wildtype protein[23].

We determined one unconventional complex splicing event, where the last 151 bp of exon 1A are skipped, exon 2A is skipped, and the first 21 bp of exon 3A are skipped. This splice variant would conform to a GC-TG splice donor-acceptor event (Appendix Figure 3), contrasting the canonical use of the GT-AG splice donor-acceptor signal nucleotides. A GC-TG splice donor-acceptor event was reported by Ikemoto et al[24] for an isoform of the gonadotropin-releasing hormone (GnRH) in pufferfish. In mammalian genes, it has been estimated that 99.24% correspond to canonical splice site (GT-AG), 0.69% corresponds to GC-AG, 0.05% to AT-AC and only 0.02% corresponds to other non-canonical splice sites[25]. Therefore, the GC-TG combination we identified is a very rare event.

In conclusion, this study is the first to characterize and report the genomic organization of the human *SLC22A23* gene, as well as the alternative splicing events leading to variable mRNA isoforms. The study will provide the basis for further functional studies and the means to identify disease mechanisms associated to variation in the *SLC22A23* gene.

6.6 REFERENCES

1. Koepsell, H. Polyspecific organic cation transporters: their functions and interactions with drugs. *Trends Pharmacol. Sci.* **25,** 375–381 (2004).
2. Tamai, I. Pharmacological and pathophysiological roles of carnitine/organic cation transporters (OCTNs: SLC22A4, SLC22A5 and Slc22a21). *Biopharm. Drug Dispos.* **34,** 29–44 (2013).
3. Shu, Y. Research progress in the organic cation transporters. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* **36,** 913–926 (2011).
4. Jacobsson, J. A., Haitina, T., Lindblom, J. & Fredriksson, R. Identification of six putative human transporters with structural similarity to the drug transporter SLC22 family. *Genomics* **90,** 595–609 (2007).
5. 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).
6. Aberg, K. *et al.* Genome-wide association study of antipsychotic-induced QTc interval prolongation. *Pharmacogenomics J.* **12,** 165–172 (2012).
7. Bennett, K. M., Liu, J., Hoelting, C. & Stoll, J. Expression and analysis of two novel rat organic cation transporter homologs, SLC22A17 and SLC22A23. *Mol. Cell. Biochem.* **352,** 143–154 (2011).
8. McEntyre, J. & Ostell, J. The NCBI Handbook. (2002). at <http://www.ncbi.nlm.nih.gov/books/NBK21101/>
9. Canese, K., Jentsch, J. & Myers, C. PubMed: The Bibliographic Database. (2003). at <http://www.ncbi.nlm.nih.gov/books/NBK21094/>

10. Pontius, J. U., Wagner, L. & Schuler, G. D. UniGene: A Unified View of the Transcriptome. (2003). at <http://www.ncbi.nlm.nih.gov/books/NBK21083/>

11. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).

12. Flicek, P. *et al.* Ensembl 2013. *Nucleic Acids Res.* **41,** D48–D55 (2012).

13. Maglott, D., Pruitt, K. & Tatusova, T. Gene: A Directory of Genes. (2011). at <http://www.ncbi.nlm.nih.gov/books/NBK21085/>

14. *The Proteomics Protocols Handbook*. at <http://www.springer.com/life+sciences/biochemistry+%26+biophysics/book/978-1-58829-343-5>

15. Di Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39,** W13–17 (2011).

16. Viklund, H. & Elofsson, A. Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci. Publ. Protein Soc.* **13,** 1908–1917 (2004).

17. Hennerdal, A. & Elofsson, A. Rapid membrane protein topology prediction. *Bioinforma. Oxf. Engl.* **27,** 1322–1323 (2011).

18. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res.* **41,** D344–347 (2013).

19. Zhao, Z. Z. *et al.* Evaluation of polymorphisms in predicted target sites for micro RNAs differentially expressed in endometriosis. *Mol. Hum. Reprod.* **17,** 92–103 (2011).

20. Holste, D. & Ohler, U. Strategies for identifying RNA splicing regulatory motifs and predicting alternative splicing events. *Plos Comput. Biol.* **4,** e21 (2008).

21. Zhang, L., Dresser, M. J., Chun, J. K., Babbitt, P. C. & Giacomini, K. M. Cloning and functional characterization of a rat renal organic cation transporter isoform (rOCT1A). *J. Biol. Chem.* **272,** 16548–16554 (1997).

22. Faustino, N. A. & Cooper, T. A. Pre-mRNA splicing and human disease. *Genes Dev.* **17,** 419–437 (2003).

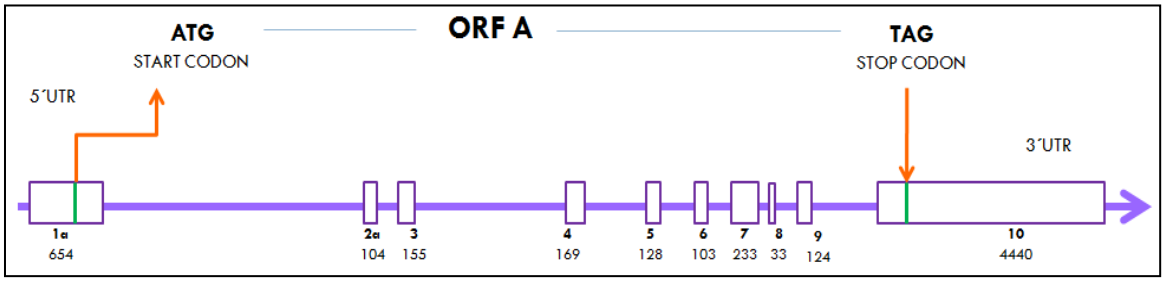
23. Maekawa, S. *et al.* OCTN2VT, a splice variant of OCTN2, does not transport carnitine because of the retention in the endoplasmic reticulum caused by insertion of 24 amino acids in the first extracellular loop of OCTN2. *Biochim. Biophys. Acta* **1773,** 1000–1006 (2007).

24. Ikemoto, T. & Park, M. K. Identification and molecular characterization of three GnRH ligands and five GnRH receptors in the spotted green pufferfish. *Mol. Cell. Endocrinol.* **242,** 67–79 (2005).

25. Burset, M., Seledtsov, I. A. & Solovyev, V. V. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res.* **28,** 4364–4375 (2000).

6.7 Figures

A



B

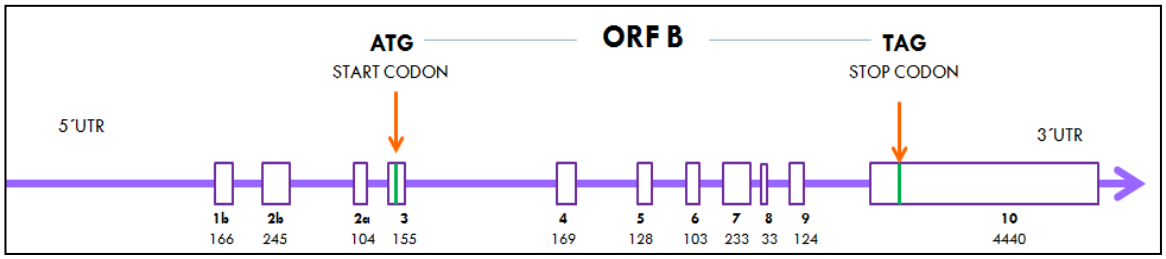


Figure 1. Genomic organization of the SLC22A23 transcripts NM015482.1 (A), and NM021945.5 (B).

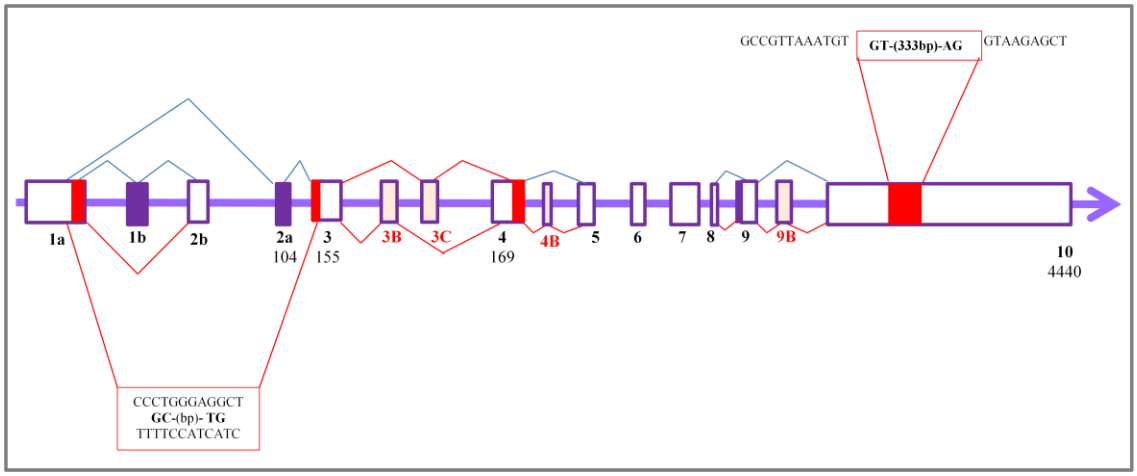


Figure 2. Alternative splicing events in the SLC22A23 locus identified based on EST alignment and visual inspection and correction.

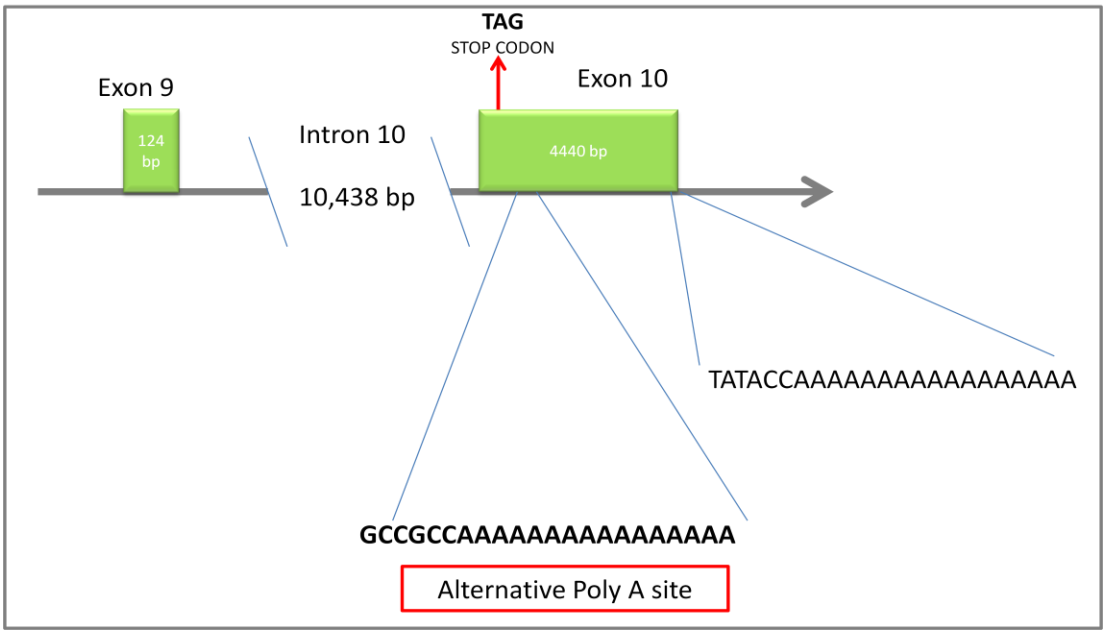
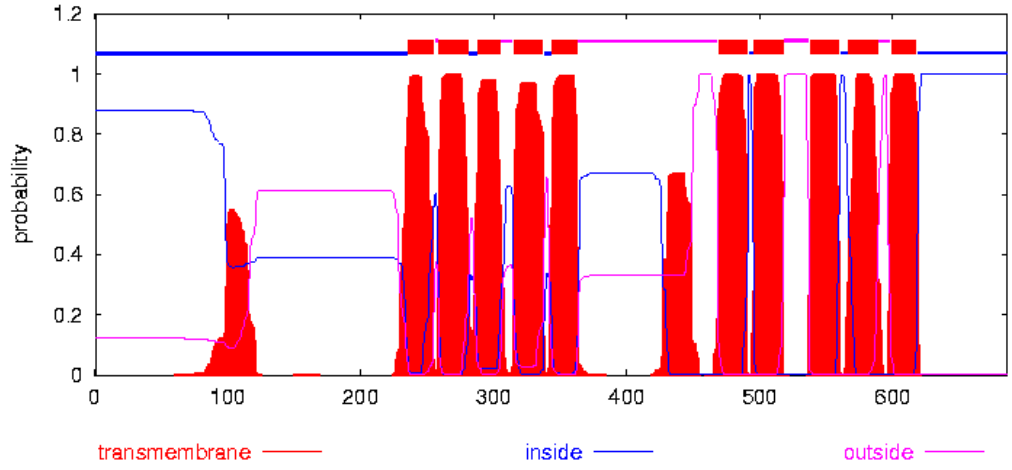


Figure 3. Alternative transcription termination sites utilized in exon 10 of the *SLC22A23* locus.

A



B

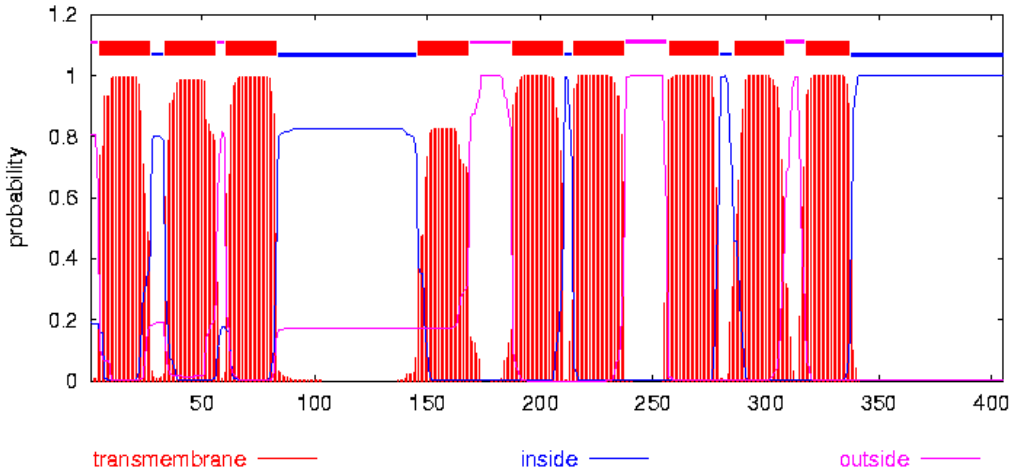


Figure 4. Predicted transmembrane architecture for the SLC22A23 protein isoform A (NP056297.1) shown in panel A, and isoform B (NP068764.3) shown in panel B. Although they only differ in their N-terminal sequence, the predicted architectures are distinct. The analysis was performed using TMHMM.

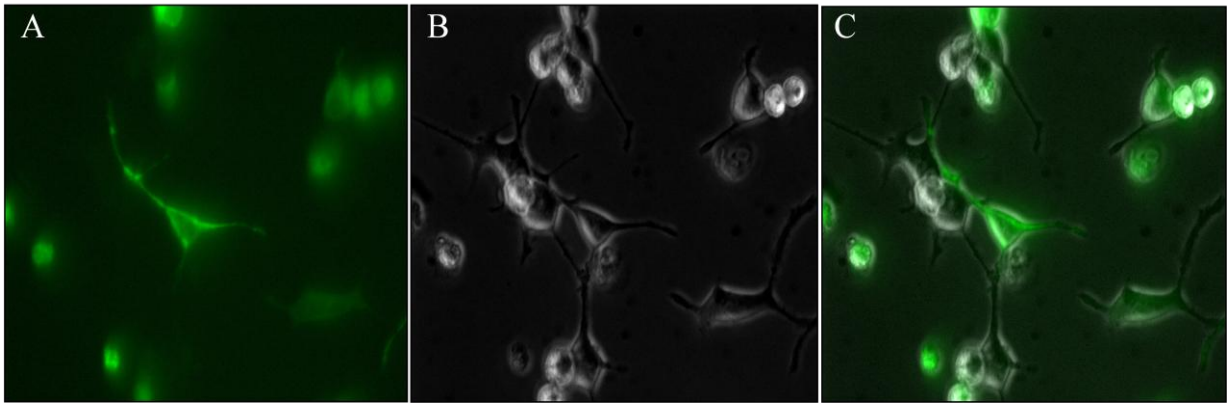


Figure 5. SLC22A23 protein expressed in cell membrane**.**The green fluorescent signal for a SLC22A23 isoform B tGFP fusion protein locates to the plasmalemma membrane in transiently transfected HEK293 cells. Panel A: GFP channel; panel B: bright light channel; panel C: merge of panel A and panel B.

6.8 Appendix

Methodology for sub-cloning human *SLC22A23* transcripts

SLC22A23 isoform A was subcloned from and isoform B was subcloned from duodenum cDNA. They were tagged with N-terminal mCherry and tGFP fluorescent protein, respectively. The Open Reading Frame (ORF) of *SLC22A23* Isoform A was amplified with sense primer 5´-ATGGGCAACTGGACCAGCCT-3´ and antisense primer 5´-CTACATGGCCTTCATGCCGT-3´.The ORF of *SLC22A23* Isoform B was amplified with sense primer 5´-ATGTTCAGCACACTCAGGTT-3´and antisense primer 5´-CTACATGGCCTTCATGCCGT-3´. The PCR amplifications were performed under the following conditions: initial denaturation at 95ºC for 30s, followed by 35 cycles of denaturation at 95ºC for 15s, annealing at 62ºC for 15s, extension at 68ºC for 2 min and final extension at 68ºC for 5 min. The reaction consisted in 20µl total volume, conformed of 1µl of forward primer (10µmol) and 1µl of reverse primer (10µmol), 4 µl of Taq 5X Master Mix (New England BioLabs), 2µl of cDNA and 12µl of dH_2O. We confirm the sequence using TOPO TA Cloning® Kit for Sequencing (Invitrogen).

We have created an construct using Invitrogen Gateway® Destination for tGFP-*SCL22A23* Isoform B with N-terminal expression using the following approach.

The fluorescent protein turbo-GFP of 699bp was subcloned from GIPZ Lentiviral shRNAmir vector (ThermoScientific Open Biosystem) with the sense primer 5´-GCCACCATGGAGAGCGACGAGAGC-3´ and antisense primer 5´-TTCTTCACC GGCATCTGC-3´ (no stop codon). Separately, we created a *SLC22A23* Isoform B PCR fragment overlapping with tGFP to be further fused in a second round PCR. We

amplified it using tGFP-R1 nostop-*SLC22A23* B forward fusion primer 5´-GCAGATGCCGGTGAAGAAATGTTCAGCACACTCAGGTT-3´ and antisense primer 5´-CTACATGGCCTTCATGCCGT-3´. The TOPO construct created with the ORF was used as template.

A second round PCR was followed to fuse tGFP to the 5´ end of *SLC22A23* Isoform B using AttB primers for Gateway Invitrogen cloning: tGFP -ATG-Kozak-Forward-AttB1 sense primer 5´-GGGGACAAGTTTGTACAAAAAAGCAGGCTTC GCCACCATGGAGAGCGACGAGAGC-3´ and *SLC22A23*-B-reverse-stop-AttB2 antisense primer 5´-GGGGACCACTTTGTACAAGAAAGCTGGGTC CTACATGGCCTTCATGCCGT-3´. PCR cycling conditions: initial denaturation at 95ºC for 30s, followed by 35 cycles of denaturation at 95ºC for 15s, annealing at 62ºC for 15s, extension at 68ºC for 2 min and final extension at 68ºC for 5 min.

The second round PCR construct (tGFP-*SLC22A23* Isoform B) was purified from the gel (QIAquick QIAGEN Gel Extraction Kit) and subcloned into pDONR™221 Gateway entry vector by BP cloning reaction, according to the manufacturers (Invitrogen)

The tGFP-*SLC22A23*-Isoform B insert was shuttle into the pcDNA/V5-DEST Gateway destination vector by Gateway LR cloning reaction. pcDNA™/V5-DEST expression vector was used to express the isoform B in HEK293 cell line. All plasmid samples were verified by sequencing at The Centre for Applied Genomics (TCAG) in Toronto.

Primer Sequences for Isoform A Cloning

Name of Primer	Sequence
SLC22A23 A F1	5´-ATGGGCAACTGGACCAGCCT-3´
SLC22A23 R1stop	5´-CTACATGGCCTTCATGCCGT-3´
SLC22A23 A R1stop attB2:	5´-GGGGACCACTTTGTACAAGAAAGCTGGGTC CTACATGGCCTTCATGCCGT-3´
mCherry-ATG-Kozak-F1:	5´-GCCACCATGGTGAGCAAGGGCGAG-3´
mCherry-ATG-Kozak-F1-attB1:	5´GGGGACAAGTTTGTACAAAAAAGCAGGCTTC GCCACCATGGTGAGCAAGGGCGAG-3´
mCherry R1 nostop:	5´-CTTGTACAGCTCGTCCAT-3´
mCherry R1 nostop-SLC22A23 A F1 phusion primer:	5´-GACGAGCTGTACAAG ATGGGCAACTGGACCAGCCT-3´

Primer Sequences for Isoform B Cloning

Name of Primer	Sequence
SLC22A23 B F1	5´-ATGTTCAGCACACTCAGGTT-3´
SLC22A23 R1stop	5´-CTACATGGCCTTCATGCCGT-3´
SLC22A23 B R1stop attB2:	5´-GGGGACCACTTTGTACAAGAAAGCTGGGTC CTACATGGCCTTCATGCCGT-3´
tGFP-ATG-Kozak-F1:	5´-GCCACCATGGAGAGCGACGAGAGC-3´
tGFP -ATG-Kozak-F1-attB1	5´-GGGGACAAGTTTGTACAAAAAAGCAGGCTTC GCCACCATGGAGAGCGACGAGAGC-3´
tGFP R1 nostop	5´-TTC TTC ACC GGC ATC TGC-3´
tGFP R1 nostop-SLC22A23 B F1 phusion primer:	5´-GCAGATGCCGGTGAAGAAATGTTCAGCACACTCAGGTT-3´

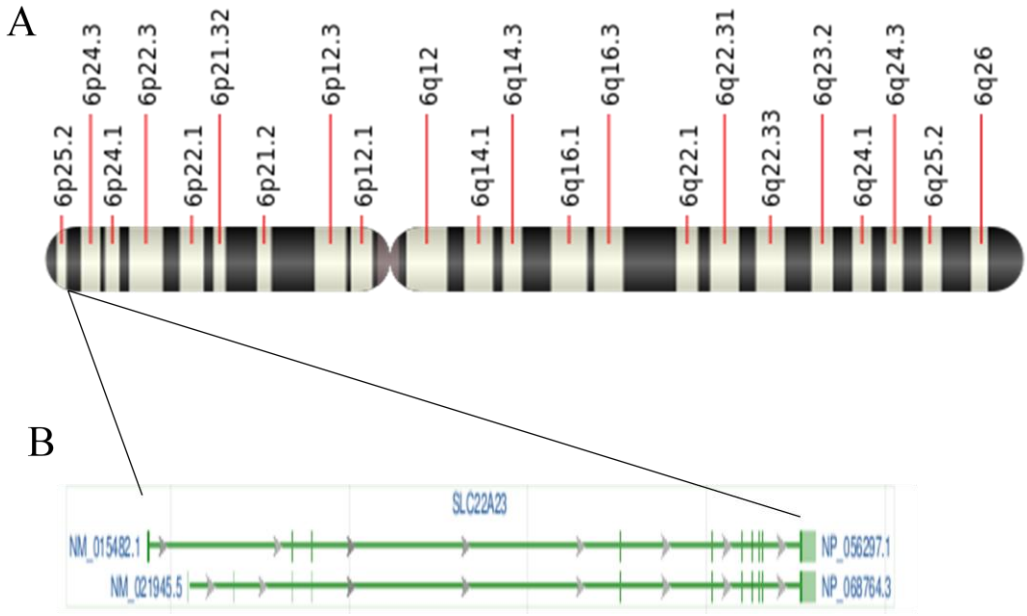


Figure 1. SLC22A23 gene localization. the *SLC22A23* gene is localized on chromosome 6p25.2, bases 3,269,196 - 3,457,256 on the reverse strand (A), and the two reference transcripts NM015482.1 (NP056297.1, isoform A) and NM021945.5 (NP068764.3, isoform B) are spanning a locus of 187.59 kb.

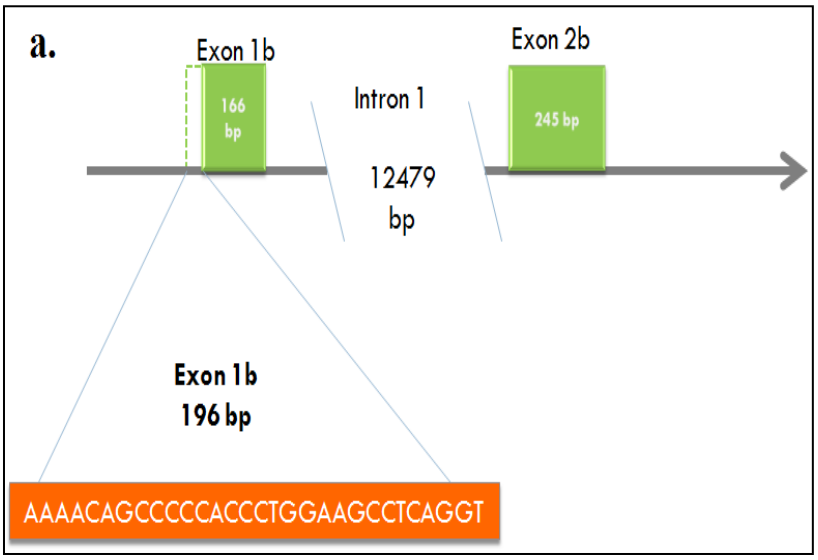


Figure 2. alternative 5' exons used by transcript NM021945.5. Currently this reference transcript for the *SLC22A23* gene is automatically annotated in gene bank to possess a first exon of 166 bp length (green box). However, evidence from EST Hs#S29618082 indicates a transcriptional start point 30 bp upstream, extending the exon to a total of 196 bp.

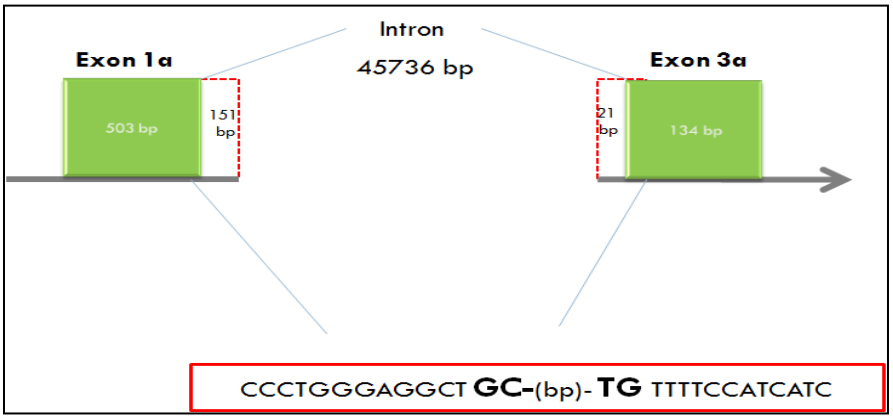


Figure 3. *SLC22A23* splicing event where the last 151 bp of exon 1A are skipped, exon 2A is skipped, and the first 21 bp of exon 3A are skipped. This splice variant would conform to a GC-TG splice donor-acceptor event.

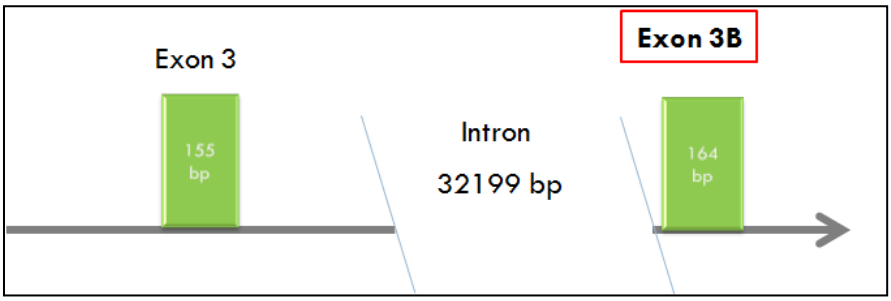


Figure 4. An alternative 164bp exon 3B could be utilized in intron 3

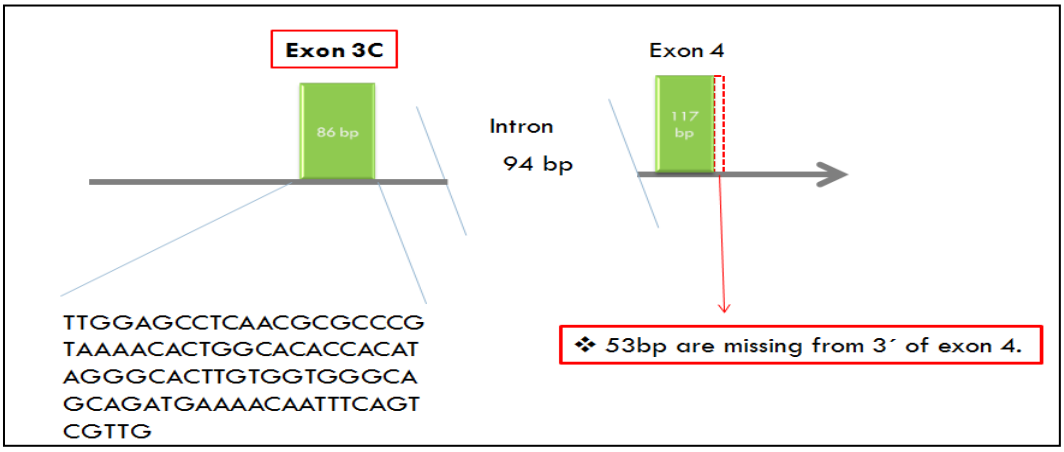


Figure 5. An alternative 86 bp exon 3C could be utilized in intron 3 and the last 53 bp of exon 4 can be skipped.

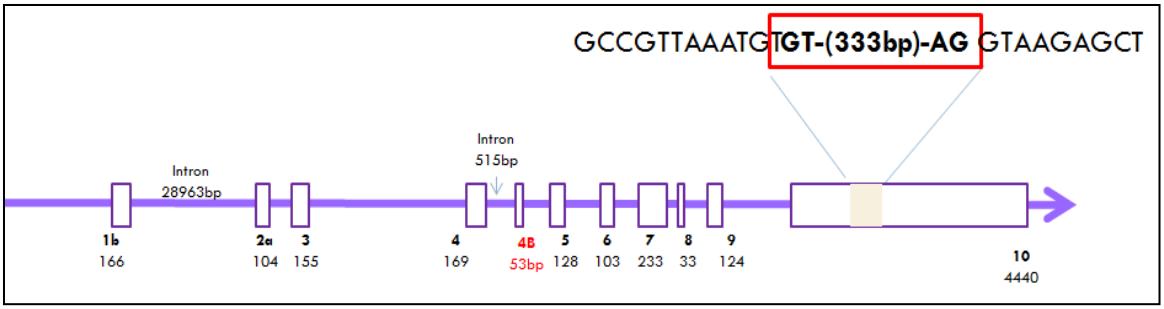


Figure 6. An alternative 53 bp exon 4B could be utilized in intron 4

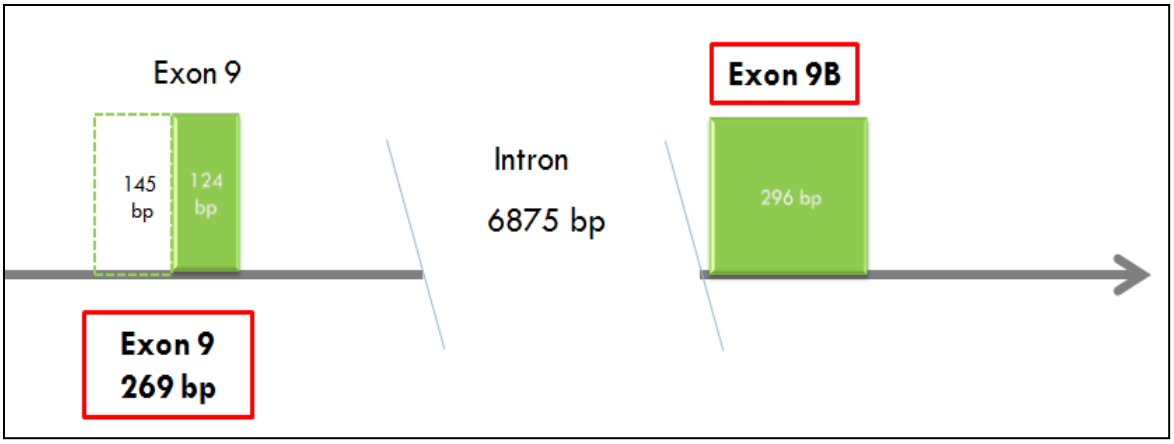


Figure 7. An alternative 296 bp exon 9B could be utilized in intron 9.

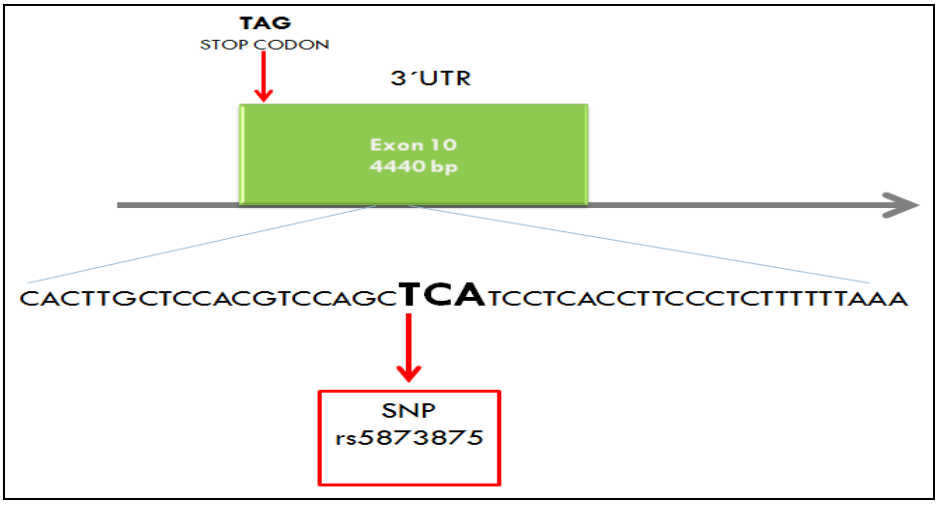


Figure 8. Insertion in 3´UTR corresponding to the SNP rs5873875.

SLC22A23 Isoform A protein sequence

MAIDRRREAAGGGPGRQPAPAEENGSLPPGDAAASAPLGGRAGPGGGAEIQPLP
PLHPGGGPHPSCCSAAAAPSLLLLDYDGSVLPFLGGLGGGYQKTLVLLTWIPALF
IGFSQFSDSFLLDQPNFWCRGAGKGTELAGVTTTGRGGDMGNWTSLPTTPFATA
PWEAAGNRSNSSGADGGDTPPLPSPPDKGDNASNCDCRAWDYGIRAGLVQNVV
SKWDLVCDNAWKVHIAKFSLLVGLIFGYLITGCIADWVGRRPVLLFSIIFILIFGLT
VALSVNVTMFSTLRFFEGFCLAGIILTLYALRIELCPPGKRFMITMVASFVAMAG
QFLMPGLAALCRDWQVLQALIICPFLLMLLYWSIFPESLRWLMATQQFESAKRLI
LHFTQKNRMNPEGDIKGVIPELEKELSRRPKKVCIVKVVGTRNLWKNIVVLCVNS
LTGYGIHHCFARSMMGHEVKVPLLENFYADYYTTASIALVSCLAMCVVVRFL
GRRGGLLLFMILTALASLLQLGLLNLIGKYSQHPDSGMSDSVKDKFSIAFSIVGMF
ASHAVGSLSVFFCAEITPTVIRCGGLGLVLASAGFGMLTAPIIELHNQKGYFLHHII
FACCTLICIICILLLPESRDQNLPENISNGEHYTRQPLLPHKKGEQPLLLTNAELKD
YSGLHDAAAAGDTLPEGATANGMKAM

SLC22A23 Isoform B protein sequence

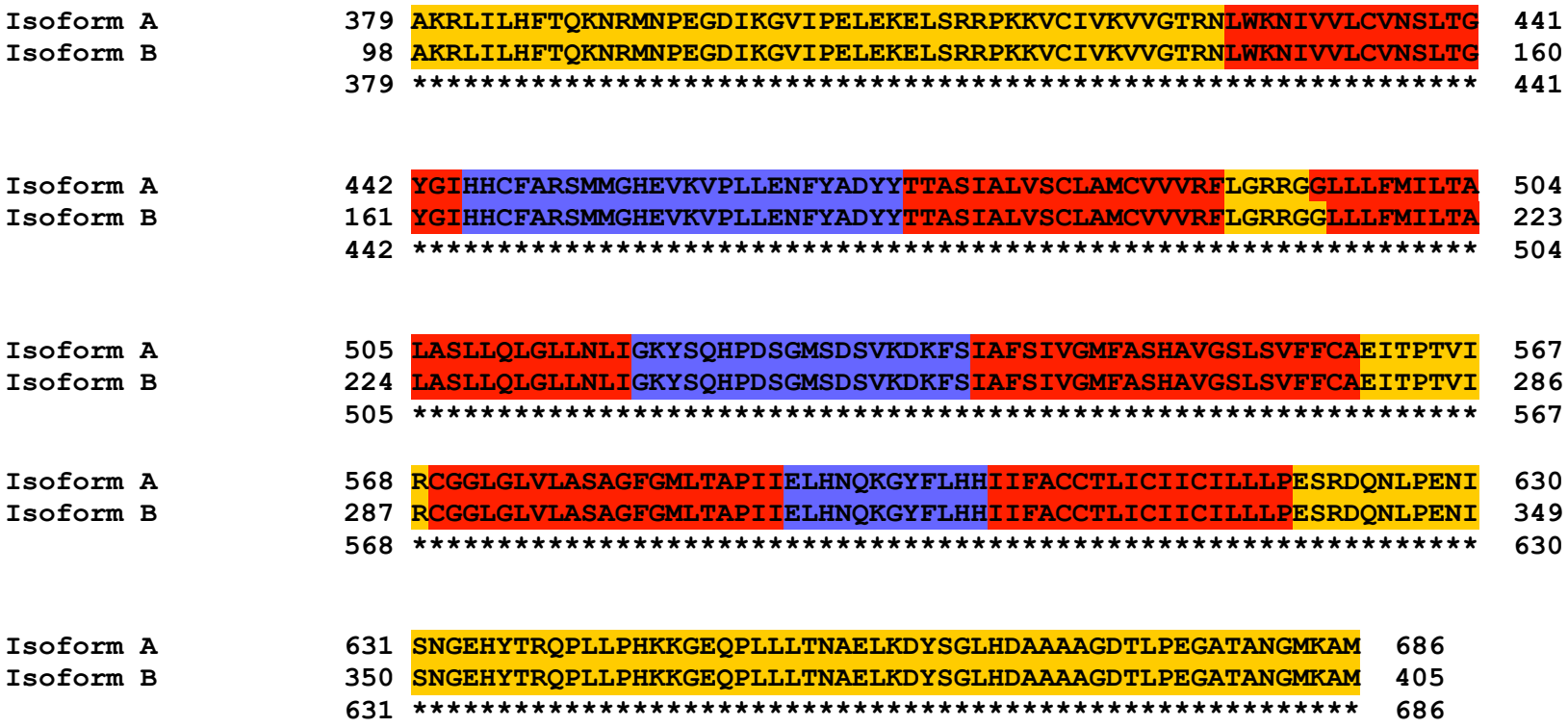
MFSTLRFFEGFCLAGIILTLYALRIELCPPGKRFMITMVASFVAMAGQFLMPGLA
ALCRDWQVLQALIICPFLLMLLYWSIFPESLRWLMATQQFESAKRLILHFTQKNR
MNPEGDIKGVIPELEKELSRRPKKVCIVKVVGTRNLWKNIVVLCVNSLTGYGIHH
CFARSMMGHEVKVPLLENFYADYYTTASIALVSCLAMCVVVRFLGRRGGLLLF
MILTALASLLQLGLLNLIGKYSQHPDSGMSDSVKDKFSIAFSIVGMFASHAVGSLS
VFFCAEITPTVIRCGGLGLVLASAGFGMLTAPIIELHNQKGYFLHHIIFACCTLICII
CILLLPESRDQNLPENISNGEHYTRQPLLPHKKGEQPLLLTNAELKDYSGLHDAA
AAGDTLPEGATANGMKAM

Alignment of Isoform A and B protein sequence

T-COFFEE, Version_9.03.r1318 (2012-07-12 19:05:45 - Revision 1318 - Build 366) Cedric Notredame

IN HEL OUT

```
Isoform A             1 MAIDRRREAAGGGPGRQPAPAEENGSLPPGDAAASAPLGGRAGPGGGAEIQPLPPLHPGGGPH   63
Isoform B             1 ------------------------------------------------------------     0
                      1                                                                 63

Isoform A            64 PSCCSAAAAPSLLLLDYDGSVLPFLGGLGGGYQKTLVLLTWIPALFIGFSQFSDSFLLDQPNF  126
Isoform B             1 ------------------------------------------------------------     0
                     64                                                                 126

Isoform A           127 WCRGAGKGTELAGVTTTGRGGDMGNWTSLPTTPFATAPWEAAGNRSNSSGADGGDTPPLPSPP  189
Isoform B             1 ------------------------------------------------------------     0
                     127                                                                 189

Isoform A           190 DKGDNASNCDCRAWDYGIRAGLVQNVVSKWDLVCDNAWKVHIAKFSLLVGLIFGYLITGCIAD  252
Isoform B             1 ------------------------------------------------------------     0
                     190                                                                 252

Isoform A           253 WVGRRPVLLFSIIFILIFGLTVALSVNVTMFSTLRFFEGFCLAGIILTLYALRIELCPPGKRF  315
Isoform B             1 ----------------------------MFSTLRFFEGFCLAGIILTLYALRIELCPPGKRF   34
                     253                            ********************************  315

Isoform A           316 MITMVASFVAMAGQFLMPGLAALCRDWQVLQALIICPFLLMLLYWSIFPESLRWLMATQQFES  378
Isoform B            35 MITMVASFVAMAGQFLMPGLAALCRDWQVLQALIICPFLLMLLYWSIFPESLRWLMATQQFES   97
                     316 ************************************************************  378
```


Isoform A 379 AKRLILHFTQKNRMNPEGDIKGVIPELEKELSRRPKKVCIVKVVGTRNLWKNIVVLCVNSLTG 441
Isoform B 98 AKRLILHFTQKNRMNPEGDIKGVIPELEKELSRRPKKVCIVKVVGTRNLWKNIVVLCVNSLTG 160
379 441
Isoform A 442 YGIHHCFARSMMGHEVKVPLLENFYADYYTTASIALVSCLAMCVVVRFLGRRGGLLLFMILTA 504
Isoform B 161 YGIHHCFARSMMGHEVKVPLLENFYADYYTTASIALVSCLAMCVVVRFLGRRGGLLLFMILTA 223
442 504
Isoform A 505 LASLLQLGLLNLIGKYSQHPDSGMSDSVKDKFSIAFSIVGMFASHAVGSLSVFFCAEITPTVI 567
Isoform B 224 LASLLQLGLLNLIGKYSQHPDSGMSDSVKDKFSIAFSIVGMFASHAVGSLSVFFCAEITPTVI 286
505 567
Isoform A 568 RCGGLGLVLASAGFGMLTAPIIELHNQKGYFLHHIIFACCTLICIICILLLPESRDQNLPENI 630
Isoform B 287 RCGGLGLVLASAGFGMLTAPIIELHNQKGYFLHHIIFACCTLICIICILLLPESRDQNLPENI 349
568 630
Isoform A 631 SNGEHYTRQPLLPHKKGEQPLLLTNAELKDYSGLHDAAAAGDTLPEGATANGMKAM 686
Isoform B 350 SNGEHYTRQPLLPHKKGEQPLLLTNAELKDYSGLHDAAAAGDTLPEGATANGMKAM 405
631 686

CHAPTER 7. SUMMARY AND GENERAL CONCLUSION

Inflammatory Bowel Disease is a disease particularly important in Canada because it has the highest incidence and prevalence rates worldwide[44]. In 2012 it was estimated that 233,000 Canadians are living with the disease[45].

The disease causing mechanism of IBD is not completely understood. Therefore, the therapeutic management has focused on the amelioration of symptoms and induction of remission, but no cure exists for the disease[46]. Current treatments (corticosteroids, aminosalicylates and immunomodulators) are associated with adverse effects. Therefore, long term safe and alternative treatments are very promising[46].

Organic cation transporters are primary targets for dietary interventions. Previously polymorphisms in *SLC22A4*, *SLC22A5* and *SLC22A23* indicate that imbalance in organic cations (metabolite or nutrient) turn the epithelial barrier more susceptible to trigger inflammation.

In this study we confirmed previously reported associations of common alleles rs17622208-A and rs11739135-C in the *SLC22A5* gene with CD, and rs11739135-C to UC in a Canadian Caucasian population. Our results suggest that polymorphisms in *SLC22A5* are involved in the development of IBD. Carnitine as its principal substrate, has been involved in the modulation and improvement of intestinal inflammation in knockout mice models[38,39] and it also has been used as an alternative therapy in clinical studies[40,47,41].

The associated SNPs are localized in intron 2 and distal of the 3´UTR of the *SLC22A5*, which indicate that they are tagging the true causal variation, lying in important protein-regulatory locus, either in the same gene or within the *IBD5* locus. The *IBD5* locus contains other immune related genes (interleukin-4 (IL4), IL13, IL5 and interferon regulatory factor-1 (IRF1)[16], therefore one might argue that the associations of organic cation transporters are in linkage with the cytokine cluster in the region.

However, previous Genome Wide Associations linked *SLC22A23*, another organic cation transport to IBD[48]. Here, for the first time, we strongly associate genotypes rs4959235-TT and rs950318-GG in the *SLC22A23* gene to UC, and potentially to IBD in a Caucasian cohort. These polymorphisms were only detected in IBD inflicted individuals increasing severely the risk for carriers of these genotypes. The localization of these SNPs do not suggest a functional impact on the protein, thus they might be linked to SNPs that have a functional effect in the same gene. Contrary to *SLC22A4* and *SLC22A5*, *SLC22A23* is located far from immune related genes, which strengthen our hypothesis that an imbalance in organic cations contribute to the development of IBD.

Further genetic studies are needed to determine the role of other variations and the functional impact on protein function. In addition, *SLC22A23* is still an orphan gene, thus the elucidation of the substrate will contribute to the understanding of the mechanism involved and will be important for further alternative therapy.

We attempted to understand the biology of the *SLC22A23*. For the first time, we subcloned and expressed the human *SLC22A23* gene and confirm the membrane localization of the protein, a fundamental requisite to function as organic cation

transporter. Our study provides the basis for further functional studies and identification of the substrate.

Overall, there is strong evidence that organic cation transporters of the SLC22A family are involved in the complex cause of IBD.

Validation of the polymorphisms are needed, however, the strong associations observed in this study can ultimately be used as predictive biomarkers for the risk of IBD. Moreover, the elucidation of the substrate of *SLC22A23* will represent an interesting target for future alternative effective dietary interventions, based on genotype. Potential nutraceuticals or functional foods can be developed avoiding the side effects of current therapies, as well as reducing the economic cost generated by those expensive treatments.

Future work directions of this study will be to validate the association of the polymorphism in different IBD populations. Moreover, we encourage further functional studies to identify the substrate of *SLC22A23*. Transport studies can be performed using *SLC22A23* stable cell line. *Slc22a23*$^{-/-}$ knock out mouse models can be developed to determine the functional role of this orphan gene.

GENERAL APPENDIX

Appendix 1. Methodology for PCR-RFLP genotyping

1. OBJECTIVE

To design new PCR-RFLP assays to genotype rs1050152, rs11568500, rs11568510, rs12521868, rs11739135, rs17622208 and rs17309827.

2. MATERIALS AND METHODS

DNA samples from IBD patients were used to identify the different genotypes. Polymorphisms in *SLC22A4* (rs1050152, rs11568500, rs11568510), in *IBD5* locus (rs12521868, rs11739135, rs17622208) and in *SLC22A23* (rs17309827) were selected to design new PCR-RFLP analysis.

First, to design the PCR-RFLP assays, the genome sequence containing the polymorphism was downloaded from NCBI data base SNP (http://www.ncbi.nlm.nih.gov/ projects/SNP/) in FASTA files. We create two sequences with either wild type or risk genotype by changing the SNP using Sequencher 4.8. The design of primers and selection of restriction enzymes was carried out by the same program.

2.1 PCR amplification

The PCR amplifications were performed in Veriti 96-well thermal cycler from Applied Biosystem. Primers used for each genotype are listed in Table 1. The following conditions were the same for all the polymorphisms: initial denaturation at 95ºC for 30s,

followed by 35 cycles of denaturation at 95ºC for 15s, annealing at 50ºC for 15s, extension at 68ºC for 2 min and final extension at 68ºC for 5 min.

Each polymerase chain reaction was carried out in 10µl total volume, conformed of 0.5µl of forward primer (10µmol) and 0.5µl of reverse primer (10µmol), 2 µl of Taq 5X Master Mix (New England BioLabs), 1µl of each DNA samples (10ng/µl) and 6µl of dH_2O.

2.2 RFLP analysis

The amplified PCR products were digested by allele-specific restriction endonucleases. The 10µl PCR product was added with 5µl of digestion reaction. The later consisted of 0.2µl of endonuclease, 1U (New England BioLabs), 1.5µl of 10X buffer and 3.3µl of dH_2O. Only RFLP analysis for rs1050152 and rs17309827 included 0.15µl of 100X BSA. Incubation was at 37ºC for 4 hours. Restriction enzymes and buffers for each SNP are detailed in Table 1.

Table 1. Primer sequence and restriction enzymes for RFLP genotyping

Gene	dbSNP	Primer Sequence 5´- 3´	PCR Size (bp)	Endonuclease	Buffer
SLC22A4	rs1050152	Forward: TTGATGTTCTTATGTCCCGG Reverse: TGTGCCCAGCCAACAATATG	309	MnlI	Buffer 4, BSA
SLC22A4	rs11568500	Forward: ACCTTGGCAACCTACACATC Reverse: TTCAGAGGGTTAGAGGGA	253	Sau96I	Buffer 4
SLC22A4	rs11568510	Forward: TTCCTTGGCAGTGGAATCTG Reverse: GAACAAAAGTGTGTCCAGGT	312	BspmI	Buffer 3
IGR2096	rs12521868	Forward: ATCCTCCATGCTACTGCT Reverse: TGGTGTAGCCAGAGTAGA	308	DraI	Buffer 4
IGR2198	rs11739135	Forward: ACTGGCTCTTTACCTGGGAA Reverse: AACTAGTCCCAACGAGATGA	369	SfaNI	Buffer 3
IGR2230	rs17622208	Forward: AGGTCTATTCCCAGGGAA Reverse: ACTCAGAAGCTGTCCATC	288	DdeI	Buffer 3
SLC22A23	rs17309827	Forward: GGAACGTACAATTCTGCA Reverse: GCATGTGAGCGTTTGATG	219	NlaIII	Buffer 4, BSA

After the digestion, 3µl was added to the samples of 6X loading dye, blue (New England BioLabs). Restriction fragments were separated by electrophoresis on 2% UltraPure agarose gel (Invitrogen) and stained with ethidium bromide for visualization with UV light (Gel Doc, BIO-RAD). 50bp DNA ladder was used to determine the product size (New England BioLabs).

2.3 Generation of TOPO plasmid-controls

Restriction digestion is considered a critical stage in the genotyping process; inaccurate digestion may give misleading results. Hence, the inclusion of quality controls is of outstanding importance. In the present genotype study, TOPO TA Cloning® Kits for Sequencing (Invitrogen) were used to establish plasmid positive controls.

PCR amplifications, followed by RFLP and electrophoresis visualization analysis of selected DNA samples lead to the determination of such samples that contain either the wild type allele or the risk allele. The sequence confirmation of selected genotypes was carried out by insertion of the sequence of interest into the plasmid.

The insert DNA must be amplified using Taq polymerase, which adds deoxyadenosine (A) to the 3´ ends of the PCR products. The (A) overhang will efficiently allow the PCR insertion into the TOPO vector, which has deoxythymidine (T) residues overhanging in the 3´ end.

Each TOPO cloning reaction included 1µl of PCR product, 0.5µl of salt solution, 0.5µl of TOPO vector and 1 µl of dH_2O in a 3µl total volume. Incubation of the reaction for 5 min at room temperature followed by transformation of the construct into competent *E.Coli* cells, according to the manufacturer (Invitrogen). *E.Coli* cells were plating in LB agar containing kanamycine and incubated overnight at 37°C.

After incubation period, some colonies were picked to analyze the insert. They were grown on LB medium containing kanamycin; the incubation period followed these conditions: overnight at 37°C and shaking at 225 rpm.

Next, plasmid isolation was carried out following the protocol for plasmid DNA purification using the QIAprep Spin Miniprep Kit and a microcentrifuge (QIAGEN). Plasmid DNA concentration was measured using NanoDrop 2000 (Thermo Scientific). All plasmid samples were sent for sequence to The Centre for Applied Genomics (TCAG) in Toronto, and confirmation was obtained for the correct sequence of the genotypes tested.

3. RESULTS

3.1 Design of PCR-RFLP assays

The primer design was carried out using the Sequencher 4.8 software. The sequence length of each primer was established to be between 18-22 base pair because is considered the optimal length to bind easily and be more specific when annealing to the template. Moreover, the GC content was considered, so no more than 50% was present in the sequences. The melting temperature (Tm) was similar for all the primers, therefore the PCR amplification was carried out using the same conditions.

The alignment of the DNA sequence containing the two possible polymorphisms using Sequencher 4.8 allowed determining which endonuclease can discriminate the variation and produce visible fragments of DNA. Primer sequences and enzymes used for each polymorphism are shown in table 1.

Each polymorphism has a particular cutting pattern, which allows the identification of the possible genotypes involved in every sample. The restriction fragments are visualized by

electrophoresis on 2% agarose gel stained with ethidium bromide. The cutting patterns for all the variations presented in this study are shown in Figure 2.

3.2 TOPO plasmid-controls

Representative samples of each genotype were used to develop a TOPO plasmid, which can be used as positive control in every PCR-RFLP genotyping assays. The sequences of every plasmid were confirmed by direct sequencing and are available on request.

The restriction digest cutting pattern of all the plasmid-controls was confirmed as shown in Figure 1.

Figure 1. Plasmid-control genotypes

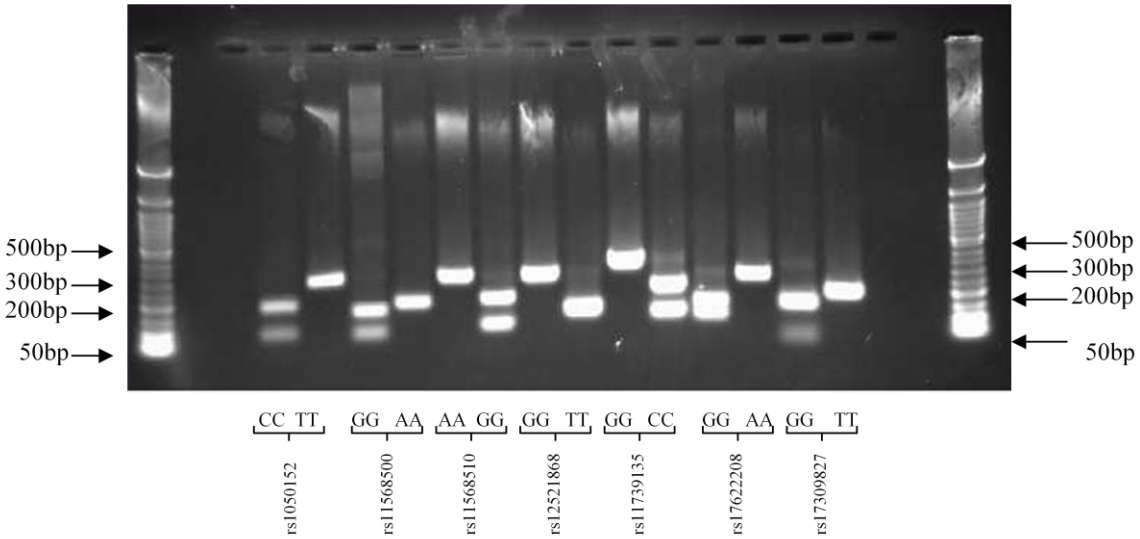
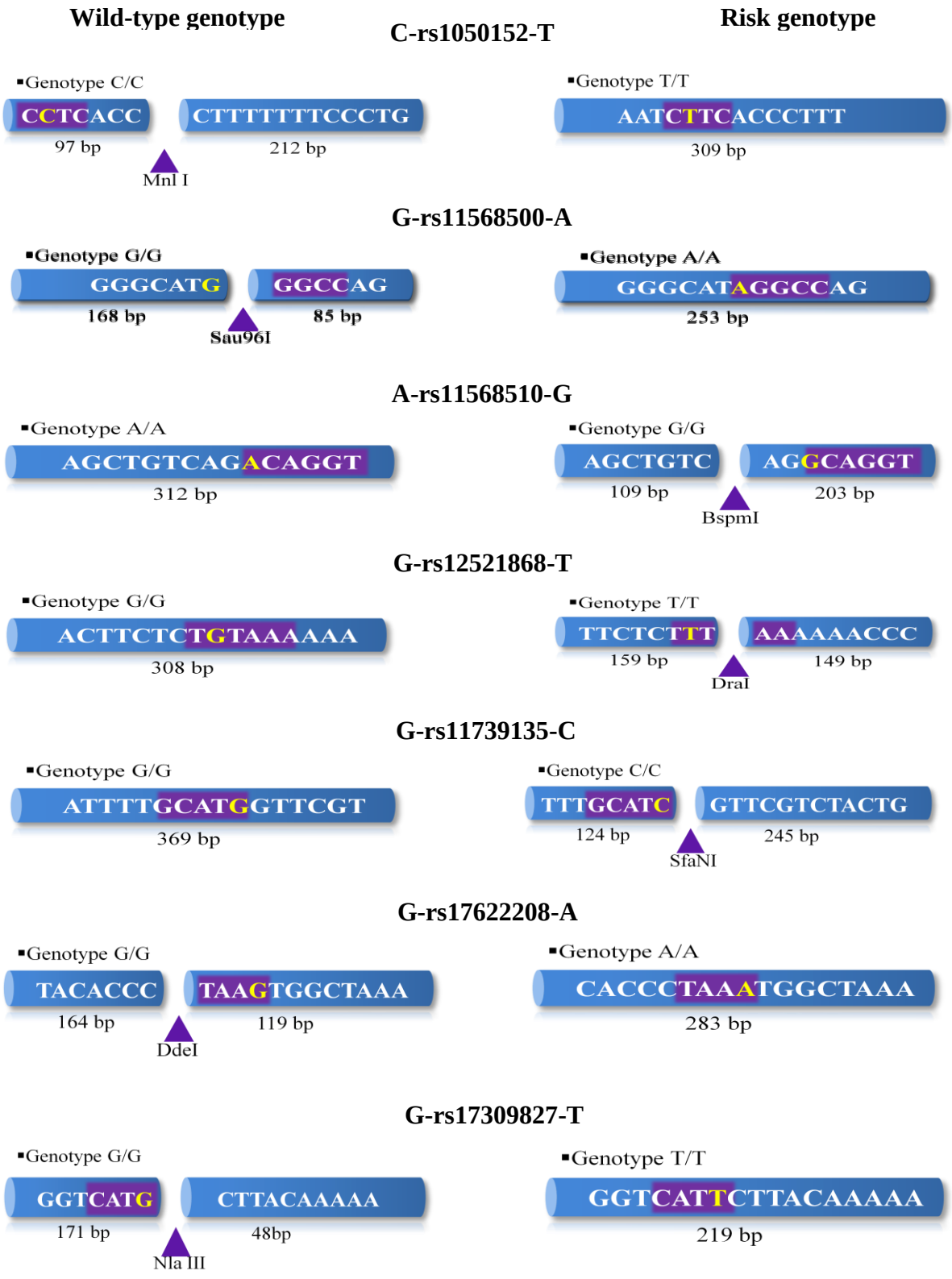


Figure 2. Restriction Digest Cutting pattern



GENERAL REFERENCES

1. Bernstein, C. N. *et al.* The epidemiology of inflammatory bowel disease in Canada: a population-based study. *Am. J. Gastroenterol.* **101,** 1559–1568 (2006).
2. Green, C., Elliott, L., Beaudoin, C. & Bernstein, C. N. A population-based ecologic study of inflammatory bowel disease: searching for etiologic clues. *Am. J. Epidemiol.* **164,** 615–623; discussion 624–628 (2006).
3. Wang, J. *et al.* Contribution of the IBD5 locus to inflammatory bowel disease: a meta-analysis. *Hum. Genet.* **129,** 597–609 (2011).
4. Xuan, C. *et al.* Association between OCTN1/2 gene polymorphisms (1672C-T, 207G-C) and susceptibility of Crohn's disease: a meta-analysis. *Int J Colorectal Dis* **27,** 11–19 (2012).
5. Tamai, I. *et al.* Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2. *J. Biol. Chem.* **273,** 20378–20382 (1998).
6. Gründemann, D. *et al.* Discovery of the ergothioneine transporter. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 5256–5261 (2005).
7. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42,** 1118–1125 (2010).
8. Fiocchi, C. Inflammatory bowel disease: etiology and pathogenesis. *Gastroenterology* **115,** 182–205 (1998).
9. Karp, L. C. & Targan, S. R. in *Inflammatory Bowel Disease* (Directoressor, S. R. T. M., Director, F. S. Md. C. & Advisor, L. C. K. R. P. S.) 3–8 (Wiley-Blackwell, 2010). at <http://onlinelibrary.wiley.com/doi/10.1002/9781444318418.ch2/summary>

10. Satsangi, J., Silverberg, M. S., Vermeire, S. & Colombel, J.-F. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* **55,** 749–753 (2006).

11. Vermeire, S., Van Assche, G. & Rutgeerts, P. Classification of inflammatory bowel disease: the old and the new. *Curr. Opin. Gastroenterol.* **28,** 321–326 (2012).

12. Levine, A. *et al.* Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflamm. Bowel Dis.* **17,** 1314–1321 (2011).

13. Podolsky, D. K. Inflammatory bowel disease. *N. Engl. J. Med.* **347,** 417–429 (2002).

14. Lees, C. W., Barrett, J. C., Parkes, M. & Satsangi, J. New IBD genetics: common pathways with other diseases. *Gut* **60,** 1739–1753 (2011).

15. Rioux, J. D. *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am. J. Hum. Genet.* **66,** 1863–1870 (2000).

16. Rioux, J. D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* **29,** 223–228 (2001).

17. Peltekova, V. D. *et al.* Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat. Genet.* **36,** 471–475 (2004).

18. Newman, B. *et al.* A risk haplotype in the Solute Carrier Family 22A4/22A5 gene cluster influences phenotypic expression of Crohn's disease. *Gastroenterology* **128,** 260–269 (2005).

19. Török, H.-P. *et al.* Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease. *Gut* **54,** 1421–1427 (2005).

20. Noble, C. L. *et al.* The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* **129,** 1854–1864 (2005).

21. Russell, R. K. *et al.* Analysis of the influence of OCTN1/2 variants within the IBD5 locus on disease susceptibility and growth indices in early onset inflammatory bowel disease. *Gut* **55,** 1114–1123 (2006).

22. Martínez, A. *et al.* Association of the organic cation transporter OCTN genes with Crohn's disease in the Spanish population. *Eur. J. Hum. Genet.* **14,** 222–226 (2006).

23. Hradsky, O. *et al.* Two independent genetic factors responsible for the associations of the IBD5 locus with Crohn's disease in the Czech population. *Inflamm. Bowel Dis.* **17,** 1523–1529 (2011).

24. Silverberg, M. S. *et al.* Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. *Eur. J. Hum. Genet.* **15,** 328–335 (2007).

25. Lakner, L. *et al.* IGR2096a_1 T and IGR2198a_1 C alleles on IBD5 locus of chromosome 5q31 region confer risk for Crohn's disease in Hungarian patients. *Int J Colorectal Dis* **24,** 503–507 (2009).

26. Repnik, K. & Potočnik, U. Haplotype in the IBD5 region is associated with refractory Crohn's disease in Slovenian patients and modulates expression of the SLC22A5 gene. *J. Gastroenterol.* **46,** 1081–1091 (2011).

27. Urban, T. J. *et al.* Functional effects of protein sequence polymorphisms in the organic cation/ergothioneine transporter OCTN1 (SLC22A4). *Pharmacogenet. Genomics* **17,** 773–782 (2007).

28. Koepsell, H. & Endou, H. The SLC22 drug transporter family. *Pflugers Arch.* **447,** 666–676 (2004).

29. Tamai, I. Pharmacological and pathophysiological roles of carnitine/organic cation transporters (OCTNs: SLC22A4, SLC22A5 and Slc22a21). *Biopharm Drug Dispos* **34,** 29–44 (2013).

30. Gründemann, D. The ergothioneine transporter controls and indicates ergothioneine activity--a review. *Prev Med* **54 Suppl,** S71–74 (2012).

31. Taubert, D., Grimberg, G., Jung, N., Rubbert, A. & Schömig, E. Functional role of the 503F variant of the organic cation transporter OCTN1 in Crohn's disease. *Gut* **54,** 1505–1506 (2005).

32. Taubert, D., Jung, N., Goeser, T. & Schömig, E. Increased ergothioneine tissue concentrations in carriers of the Crohn's disease risk-associated 503F variant of the organic cation transporter OCTN1. *Gut* **58,** 312–314 (2009).

33. Grigat, S. *et al.* Probing the substrate specificity of the ergothioneine transporter with methimazole, hercynine, and organic cations. *Biochem. Pharmacol.* **74,** 309–316 (2007).

34. Kato, Y. *et al.* Gene knockout and metabolome analysis of carnitine/organic cation transporter OCTN1. *Pharm. Res.* **27,** 832–840 (2010).

35. Petermann, I. *et al.* Mushroom intolerance: a novel diet-gene interaction in Crohn's disease. *Br. J. Nutr.* **102,** 506–508 (2009).

36. Pochini, L. *et al.* The human OCTN1 (SLC22A4) reconstituted in liposomes catalyzes acetylcholine transport which is defective in the mutant L503F associated to the Crohn's disease. *Biochim. Biophys. Acta* **1818,** 559–565 (2012).

37. Nezu, J. *et al.* Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter. *Nat Genet* **21,** 91–94 (1999).

38. Shekhawat, P. S. *et al.* Spontaneous development of intestinal and colonic atrophy and inflammation in the carnitine-deficient jvs (OCTN2(-/-)) mice. *Mol. Genet. Metab.* **92,** 315–324 (2007).

39. Sonne, S., Shekhawat, P. S., Matern, D., Ganapathy, V. & Ignatowicz, L. Carnitine deficiency in OCTN2-/- newborn mice leads to a severe gut and immune phenotype with widespread atrophy, apoptosis and a pro-inflammatory response. *PLoS ONE* **7,** e47729 (2012).

40. Fortin, G. *et al.* L-carnitine, a diet component and organic cation transporter OCTN ligand, displays immunosuppressive properties and abrogates intestinal inflammation. *Clin. Exp. Immunol.* **156,** 161–171 (2009).

41. Mikhailova, T. L. *et al.* Randomised clinical trial: the efficacy and safety of propionyl-L-carnitine therapy in patients with ulcerative colitis receiving stable oral treatment. *Aliment. Pharmacol. Ther.* **34,** 1088–1097 (2011).

42. Jacobsson, J. A., Haitina, T., Lindblom, J. & Fredriksson, R. Identification of six putative human transporters with structural similarity to the drug transporter SLC22 family. *Genomics* **90,** 595–609 (2007).

43. Bennett, K. M., Liu, J., Hoelting, C. & Stoll, J. Expression and analysis of two novel rat organic cation transporter homologs, SLC22A17 and SLC22A23. *Mol. Cell. Biochem.* **352,** 143–154 (2011).

44. Molodecky, N. A. *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* **142,** 46–54.e42; quiz e30 (2012).

45. Rocchi, A. *et al.* Inflammatory bowel disease: a Canadian burden of illness review. *Can. J. Gastroenterol.* **26,** 811–817 (2012).

46. Viladomiu, M., Hontecillas, R., Yuan, L., Lu, P. & Bassaganya-Riera, J. Nutritional protective mechanisms against gut inflammation. *J. Nutr. Biochem.* **24,** 929–939 (2013).

47. D'Argenio, G. *et al.* Experimental colitis: decreased Octn2 and Atb0+ expression in rat colonocytes induces carnitine depletion that is reversible by carnitine-loaded liposomes. *FASEB J.* **20,** 2544–2546 (2006).

48. 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).