Genetics And Disease Associations Of Organic Cation Transporters With IBD – Special Emphasis On Genetic And Functional Studies Of *SLC22A23*

By

Nazia Akhter Chaity

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Human Nutritional Sciences

University of Manitoba

Winnipeg

ABSTRACT

Inflammatory bowel disease (IBD) is a chronic disease which steadily increases worldwide with the highest prevalence in Canada. Genetic susceptibility is considered to be an important factor in causing IBD.

Organic cation transporters, *SLC22A4* and *SLC22A5* have been associated to IBD multiple times. Recently, *SLC22A23*, a novel gene that encodes for an organic cation membrane transporter protein has also been associated to IBD however; neither its gene structure nor its functions has been characterized.

The aim of this study was to characterize the genomic structure of *SLC22A23* gene using bioinformatics analysis, determine the tissue expression, characterize the location of the protein and perform functional studies using Liquid Chromatography-Quadrupole Time of Flight-Mass Spectrometry.

We have identified the chromosomal location, the gene neighborhood and the genomic structure of human *SLC22A23*.The result of this study indicates that *SLC22A23* gene is a membrane transporter and it is abundantly expressed in the intestine.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and admiration to my supervisor Dr Peter Eck, for giving me the opportunity to work in your supervision. Thank you for the confidence, guidance, inspiration, invaluable knowledge and unconditional support throughout my Master´s program. Without him completion of my master's program would not have been possible.

I extend my appreciation to my committee members Dr. William L. Diehl-Jones and Dr Michel Aliani for their guidance and contribution in this project.

I would like to thank Faculty of Graduate Studies and the Human Nutritional Science department for the financial support. My sincere appreciation to Dr Christine Zhang for helping me in the FACS method and Andre Dufresne for helping me in the fluorescence microscopy. I would also like to thank my entire laboratory group members for their support.

Finally, I would like to thank my beloved parents A K M Azizul Haque and Nilufar Yeasmin, and my elder brother Nafiz Chisty for all their blessings. Thank you for motivating and encouraging me to succeed and for all the support and love during this dream.

And my most sincere appreciation to my strength; my husband Abdullah Khafi for his immense support, inspiration and guidance. Without whom this dream would not have been possible.

DEDICATION

This thesis is dedicated to my husband and my parents for their belief and support throughout this journey. Without my family I would not be able to chase my dream.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF COPYRIGHTED MATERIAL

1. INFLAMMATORY BOWEL DISEASE

Inflammatory bowel disease (IBD) is a group of disorders that cause sections of the gastrointestinal tract to become severely inflamed and ulcerated (Pedersen et al., 2014).

Crohn's disease (CD) and ulcerative colitis (UC) are the two main subtypes of IBD (Baumgart & Carding, 2007). Crohn´s disease is characterized by inflammation at any level of the gastrointestinal tract, from mouth to anus, although it most commonly affects the small intestine (ileum) (Matricon, Barnich & Ardid, 2010), and ulcerative colitis is characterized by exclusive inflammation of mucosa at the colon level, in a continuous way and is limited to the inner lining of the intestinal wall (Matricon, Barnich & Ardid, 2010).

Being a chronic disease IBD is steadily increasing worldwide with the highest prevalence in Canada (Bernstein et al., 2006). It manifests as significant public health problem in Canada putting a sizeable burden on the health care budget (Bernstein et al., 2012). Moreover, due to intestinal inflammation, the quality of lives of many Canadians is significantly reduced and incidences are still rising – currently at 0.67% per year (Bernstein et al., 2012). In 1999 Manitoba had the highest documented rates of both Crohn's disease and ulcerative colitis in the world (Bernstein et al, 1999). According to 2012 Canadian IBD Epidemiology Database study, Ontario and Quebec has the highest prevalence of inflammatory bowel disease (Rocchi et al., 2012).

2. ECONOMIC IMPACT OF IBD

According to Crohn's and Colitis Foundation of Canada (CCFC), approximately 233,000 Canadians are living with IBD - 129,000 with Crohn's disease and 104,000 with ulcerative colitis (CCFC Report, 2012). Although it may occur at any age, the peak age of onset for IBD is 15 to 30 years old (CCFC Report, 2012). The incidence of IBD in children under the age of 10 has been rising significantly since 2001 and an estimated 5,900 Canadian children have IBD (CCFC Report, 2012). The economic costs of IBD are conservatively estimated as $2.8 billion per year, which is more than $11,900 per person every year (CCFC Report, 2012). Ulcerative colitis is slightly more common in males, whereas Crohn's disease is marginally more frequent in women (CCFC Report, 2012).

3. ETIOLOGY OF IBD

The mechanisms or factors responsible for triggering IBD are still not clear. However, IBD is hypothesized to be a multifactorial condition where complex interplay of luminal, environmental, and genetic factors trigger an inappropriate mucosal immune response (Matricon, Barnich & Ardid, 2010). Normally immune system protects the body from infection. However in people with IBD, the immune system confuses food, bacteria, and other materials in the intestine for foreign substances and it attacks the cells of the intestinal layers. This inappropriate reaction leads to chronic inflammation, ulceration, and thickening of the intestinal wall (CCFC Report, 2012).

3.1. Environmental Factors

Growing evidence suggests there are some intrinsic etiologic factors such as familial aggregation and ethnic, racial variations and extrinsic environmental etiologic factors play an important role in the incidence of IBD (Bernstein et al, 1999). Some researchers have also pointed to geographic variations in the incidence of Crohn's disease and ulcerative colitis as evidence of environmental etiologic factors (Calkins and Mendeloff, 1986; Lashner, 1995).

The highest incidence rates and prevalence of ulcerative colitis and Crohn's disease have been reported from northern Europe, the UK, and North America, and rates significantly increasing in low-incidence areas for instance southern Europe, Asia, and most developing countries(Bernstein et al., 1999; Binder et al., 1982; Blanchard et al., 2001; Daniel and Simon, 2007; Langholz et al., 1991; Lapidus, A., 2006; Loftus et al., 2006; Munkholm et al., 1992; Sedlack et al., 1980).

IBD is more prevalent in some ethnic groups, for instance Jews of European descent (Brant et al., 2007) whereas aboriginal Canadians, Americans of Hispanic, New Zealand's Maori and Asian people have lower occurrence of IBD (Blanchard et al., 2001; Eason et al., 1982;

Kurata et al., 1992; Sonnenberg et al., 1991). In North America, Caucasians and African-American people have higher rate of IBD (Kurata et al., 1992).

The reasons for these geographic variations are still unknown however they may result from variations in the geographic distribution of important environmental etiologic agents (Blanchard et al., 2001).

Some factors in the environment have been linked to IBD. Urbanization of societies, changes in diet, antibiotic use, hygiene status, microbial exposures, pollution and smoking have been implicated as potential environmental risk factors for IBD (Ng et al., 2013). Emergence of disease in developing countries suggests that epidemiological evolution is connected to westernization of lifestyle and industrialization. Since IBD is more common in developed nations, air pollution and certain pollutants were frequently associated with an increased risk of development of IBD, but only in younger patients (Kaplan et al., 2010). Diet has also been linked with IBD development, with high intake of total fat, polyunsaturated fatty acids, omega-6 fatty acids and red meat reported as increasing the risk of IBD (Hou et al., 2011). Canadian researchers have found that children who had one or more prescriptions for an antibiotic in the first year of life are more likely to develop IBD (Shaw et al., 2010).

The strongest environmental relations recognized are smoking and appendectomy (Ng et al., 2013). It has been found that ulcerative colitis is more common among ex-smokers and nonsmokers, whereas Crohn's disease is more common among smokers (Loftus, 2004) and people who have had an appendectomy have a lower occurrence of ulcerative colitis (Andersson et al., 2001). However, all these factors do not necessarily contribute in the occurrence of IBD (CCFC Report, 2012).

3.2. Genetic Susceptibility

Significant variation in the incidence and prevalence rates of IBD between different populations suggests that genetic factors play an important role in the pathogenesis of IBD (Blanchard et al., 2001).

Genetics is a major factor as there are many growing evidence. Mostly intestinal genes and immune genes are shown to be associated with IBD (Figure 1). However, presently more number of genes has been associated with IBD. The greatest relative risk of IBD disease has been found among first-degree relatives, suggesting a strong genetic component (Ek Weronica et al., 2014).

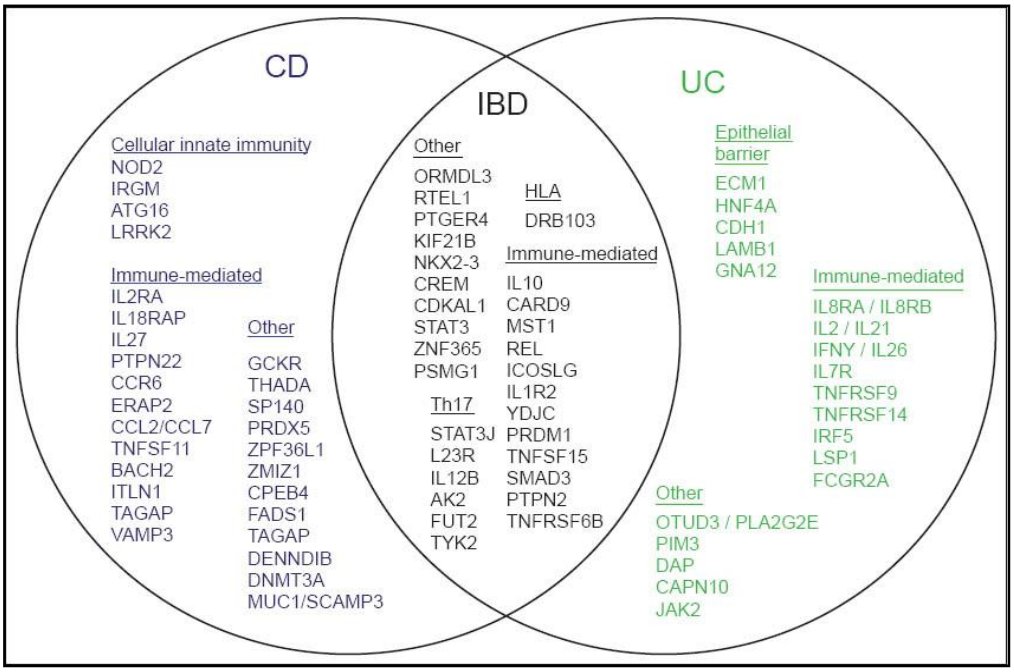


Figure 1: Inflammatory bowel disease (IBD) loci, represented by lead gene name, according to pathway. Loci associated with inflammatory bowel disease are shown in black, Crohn's disease (CD) in blue and ulcerative colitis (UC) in green (Ek et al., 2014).

The familial nature of IBD was first identified in 1909 (Allchin, 1909; Kirsner and Spencer, 1963). Aggregation of IBD cases in families has been broadly confirmed since then, with 5-

23% of patients with IBD having an affected first-degree relative (Ek Weronica et al., 2014). The risk of IBD in children increases dramatically if both parents have IBD. The estimated risk of IBD in the offspring is 33-52%, depending on follow up (Bennett et al., 1991; Laharie et al., 2001).

Further evidence for the importance of genetics in IBD comes from studies of siblings and identical twins (Blanchard et al., 2001). Siblings are most prone to be affected; the risk of IBD in a sibling is 10 to 20 times higher than the normal population (Binder and Orholm, 1996; Monsen et al., 1991). The strongest confirmation comes from twin studies where up to 50% of identical twins are most likely to have CD, while 10% will both have UC (Halfvarson et al., 2003; Orholm et al., 2000). Conversely, the opposite is also true, for at least 50% of identical twins, only one of the two will have IBD (CCFC Report, 2012). Prevalence of IBD is higher among monozygotic twins compared to dizygotic twins (Podolsky, 2002).

Although genetics is clearly a factor, the association is not simple. It is likely that more than one gene is responsible. However, just having the genes associated with IBD doesn't absolutely predict that the disease will occur. These genes are known as susceptibility genes as they enhance the chances for getting the disease. It is clear that other factors, including environmental factors, must also come into play (Facts about IBD, 2011).

Numerous genetic studies, including candidate gene approaches, linkage mapping studies, and in particular genome-wide association studies (GWASs), have allowed a better understanding and contribution of the genetic factors associated to IBD (Aberg et al., 2012; Barrett et al., 2008; McGovern et al., 2010; Rioux et al.,2000). The latest international collaborative studies have brought the number of IBD susceptibility gene loci to 163. Of the 163 identified loci, 110 are associated with both diseases, 30 are CD specific, and 23 are UC specific (Jostins et al., 2012).

Majority of the genes associated with IBD are immune genes (Ek Weronica et al., 2014; Jostins et al., 2012). Beside immune genes IBD have also been frequently associated to genetic mutations of organic cation transporters at different levels (Newman and Katherine, 2005; Peltekova et al., 2004).

4. LITERATURE REVIEWON SOLUTE CARRIERS AND IBD

4.1. Solute Carrier (SLC) Transporters

It is well known that membrane transporters are proteins that control the up-take and efflux of various solutes, including metabolites, ions, toxins, and drugs (Schlessinger et al., 2013).

Sequencing of multiple genomes, as well as expressed sequence tag (EST) approaches, have advanced the knowledge on membrane transporters further and recently facilitated the identification and characterization of many additional "missing" transporters and transporter families (Hediger et al., 2004). At present the transporter field has undergone a renaissance, and numerous transporters of various classes have since been identified (Hediger et al., 2004). They include solute carrier (SLC) transporters as well as ion dependent pumps and ATP-binding cassette (ABC) transporters (Hediger et al., 2004).

The solute carrier (SLC) transporters ranks among the largest families of membrane proteins in human which has now grown to include 52 families and 395 transporter genes in the human genome (Hediger et al., 2004). The majority of the SLC transporters are secondary active transporters, such as exchangers, symporters and antiporters, for which transport is driven by various energy coupling mechanisms (Schlessinger et al., 2013). The SLC members play an important role in a variety of cellular functions, often cooperatively with other protein families, including receptors, enzymes, and other transporters (Schlessinger et al., 2013).

4.2. Solute Carrier Family 22

Among the SLC members, solute carrier family 22 (SLC22) is a large family of organic ion transporters that belongs to the major facilitator superfamily (MFS) (Jacobsson et al., 2007).

The MFS is one of the two largest families of membrane transporters comprising uniporters, symporters and antiporters from mammals, lower eukaryotes, bacteria, and plants (Jacobsson et al., 2007).

Today it is known that members of the SLC22 family are important for the uptake, reabsorption and excretion of different compounds including drugs, environmental toxins, and endogenous metabolites across the cell membrane (Bennett et al., 2011).

Many SLC22 genes are expressed in liver, kidneys and intestine where they play important roles in absorption and excretion (Koepsell and Endou, 2004). They also perform homeostatic functions in brain and heart (Koepsell and Endou, 2004). Defect mutations of transporters of the SLC22 family may cause specific diseases such as "primary systemic carnitine deficiency" or "idiopathic renal hypouricemia" or change drug absorption or excretion (Koepsell and Endou, 2004).

The SLC22 proteins have been predicted to have 12 alpha-helical trans-membrane domains (TMDs), one extended extracellular loop between the first and the second transmembrane regions, and one extended intracellular loop between the sixth and the seventh transmembrane regions (Koepsell, 2013).

The large extracellular loop is glycosylated and mediates homo-oligomerization whereas the large intracellular loop is involved in post transcriptional regulation (Koepsell, 2013). Most transporters of the SLC22 family are considered polyspecific, i.e., they transport multiple structurally different substrates of variety of sizes and molecular structures and numerous additional compounds can act as high and/or low affinity inhibitors (Koepsell, 2013).

The family can be divided into three sub-groups according to substrate specificity and function: the organic cation transporters (OCTs), the organic cation/carnitine transporters (OCTNs), and the organic anion transporters (OATs) (Koepsell, 2013). Moreover, there are

several members of the SLC22 family with unidentified substrate specificity or function (Koepsell, 2013). OCTs function as uniporters that mediate facilitated diffusion in either direction and are responsible for translocation of (i) organic cations, such as tetraethylammonium (TEA) and 1-methyl-4-phenylpyridinium(MPP), (ii) endogenous amines, such as dopamine, (iii) therapeutic drugs, such as cimetidine and morphine, and (iv) cationic xenobiotics, such as antihistamines (Jacobsson et al., 2007). There are three distinct OCT transporters, namely, OCT1, OCT2, and OCT3, which mediate the entry of organic cations into cells (Koepsell et al., 2007).

Over the last 15 years, a number of transporters that translocate organic cations were characterized functionally and also identified on the molecular level but the human and rodent genomes have a number of SLC22 orphan genes whose functions remain unknown (Bennett et al., 2011).

As mentioned earlier, apart from immune genes, IBD have also been associated to membrane transporters particularly from the SLC family which includes*SLC22A23* (Barrett et al. 2008; Franke et al., 2010; Serrano et al., 2014).

4.3. Novel Organic Cation Transporters (OCTN) and IBD

Inflammatory bowel diseases (IBDs) have been quite clearly associated, with distinct techniques, to genetic variations of novel organic cation transporters (OCTN) (Newman and Katherine, 2005; Peltekova et al., 2004). Variation in the *SLC22A4* and *SLC22A5* genes, encoding organic cation transporters, has repeatedly been associated to IBD (Koepsell, 2013). However, studies in cohort of different regional origin and ethnic backgrounds show substantial discrepancies in disease associations (Newman & Katherine, 2005; Peltekova et al., 2004; Rioux, J. D. et al., 2000).

SLC22A4 (solute carrier family 22, member 4) is a human gene located in chromosome 5q31.1 (Koepsell, 2013). The functional protein encoded by the gene is called OCTN1 (organic cation transporter, novel, type 1) (Tamai et al., 1997).In 1997, human OCTN1 was cloned from a human fetal liver cDNA library, and rat and mouse isoforms were also subsequently isolated (Tamai et al., 1997). The protein locates to the plasma membrane; however its localization in mitochondria has also been suggested (Lamhonwah et al., 2006; Koepsell et al., 2007). It is located in the brush-border membrane of renal proximal tubule cells (Koepsell et al., 2007) and in the luminal membrane of airway epithelia (Horvath et al., 2007). In human heart OCTN1 was localized to endothelial cells of microvessels (Iwata et al., 2008).

In the first publication, OCTN1 (novel organic cation transporter) was reported to transport tetraethylammonium (TEA) and was suggested to operate as a proton/organic cation antiporter at the renal apical membrane and other tissues (Tamai et al., 1997). Later, OCTN1 was functionally demonstrated to be a multispecific and pH-dependent organic cation transporter. In the plasma membrane OCTN1 may operate as organic cation/proton exchanger or as sodium-independent transporter for zwitterions (Koepsell et al., 2007; Urban et al., 2007). OCTN1 contains an intracellular nucleotide binding site which is involved in the regulation of transport activity (Koepsell, 2013; Pochini et al., 2011; Tamai et al., 1997). Cations that are transported across the plasma membrane by human SLC22A4 are TEA, acetylcholine (Pochini et al., 2011), quinidine, verapamil, pyrilamine, the anticholinergic drugs ipratropium, tiotropium (Nakamura et al., 2010), Anticancer drugs mitoxantrone and doxorubicine and anticonvulsant drug gabapentin have been demonstrated to be substrates of the multispecific transporter (Okabe et al., 2008; Urban et al., 2007).

Human SLC22A4 also transports zwitterions like glycine-betaine, betonicine, stachydrine, ergothioneine and carnitine (Grundemann et al., 2005; Koepsell et al., 2007; Yabuuchi et al.,

1999). SLC22A4 has a low affinity to carnitine and shows high affinity to the antioxidant ergothioneine and therefore today it is considered as the principal transporter of ergothioneine (ET) (Grigat et al., 2007; Grundemann et al., 2005; Koepsell et al., 2007).

In addition, OCTN1 is believed to mediate the absorption of ergothioneine in small intestine and the reabsorption of ergothioneine in the kidney. The antioxidant ergothioneine is present in plants and mushrooms and reaches high concentrations in bone marrow and erythrocytes (Koepsell et al., 2007).

SLC22A4 is widely expressed in human body (Koepsell et al., 2007). It is expressed in kidney, ileum, colon (Meier et al., 2007), spleen, heart, skeletal muscle, brain, mammary gland (Gilchrist and Alcorn, 2010; Lamhonwah et al., 2011), lung (Horvath et al., 2007), thymus, prostate, airways, testis, bone marrow, skin (Markova et al., 2009), cornea, blood–retina barrier, iris–ciliary body (Garrett et al., 2008 & Zhang et al., 2008), fetal liver, inflammatory joints (Tokuhiroet al., 2003), sperm, placenta, uterus, immune cells, and tumor cells (Koepsell et al., 2007 & Wang et al., 2007). *SLC22A4* expression has also been described in different cancer cell lines (Okabe et al., 2008; Tamai et al., 1997).

Primarily in 2004, Peltekova et al. genotyped 469 Crohn disease individuals of European origin, resequenced 5 known genes in the IBD5 region and identified 10 new single nucleotide polymorphisms (SNPs), that includes 2 novel SNPs in organic cation transporter, (*SLC22A4* and *SLC22A5,* encoding OCTN1 and OCTN2, respectively) (Peltekova et al., 2004).

SLC22A4 (1672C or 1672T), rs1050152, cDNAs was transfected into HeLa, GM 10665 or HEK293 fibroblasts cells for carnitine or TEA uptake evaluation. Wholemount in situ hybridization and gel shift assays were used to visualizing the location of expressed RNAs and to determine the functional variants (Peltekova et al., 2004).

The result of the study showed, 2 functional mutations create a 2-allele risk haplotype TC, which had a frequency of 54% in patients with CD compared with 42% in controls, was reported to be associated with CD patients who lacked the extended IBD5 risk haplotype (Peltekova et al., 2004).

The risk for the disease was higher in the presence of the OCTN-TC, rs1050152, haplotype and at least 1 NOD2/CARD15 variant (Peltekova et al., 2004).No other studies have replicated this finding (Peltekova et al., 2004).

Amongst the 2 functional mutations, the first is a C→T substitution located in exon 9 of the *SLC22A4* gene, which results in phenylalanine replacement of the leucine residue in OCTN1 codon 503 (L503F), rs1050152.The second variation disrupts a heat shock element 207 base pairs upstream of the OCTN2 translation initiation codon (Newman et al., 2005; Peltekova et al., 2004). These alterations both have deleterious biological effects, with the 503F substitution engendering marked changes in OCTN1 transporter activity and the-207C (rs2631367) variant impairing OCTN2 promoter functions (Newman et al., 2005; Peltekova et al., 2004). The substitution impairs the carnitine transporter activity of OCTN1 but increases OCTN activity for tetraethyl ammonium and some xenobiotics (Peltekova et al., 2004). This study also reported that the two organic cation transporter genes OCTN1 (*SLC22A4*) and OCTN2 (*SLC22A5*) are located centrally on IBD5 and are independently associated with CD (Peltekova et al., 2004).

The study of Peltekova et al. has been widely replicated in different populations; however, the results from some studies have shown no significant association (Babusukumar et al., 2006; Leung et al., 2006; Lin et al., 2010).

To determine whether the previously mentioned *SLC22A4*/*SLC22A5* susceptibility alleles have impact on IBD disease location, ages at onset, or other facets of clinical presentation,

like the CD-associated *CARD15* alleles; Newman et al., examined the contribution of this susceptibility haplotype alone and in combination with *CARD15* variants to CD subphenotypes and to susceptibility to ulcerative colitis (UC) (Newman et al., 2005).

Phenotype-genotype associations were evaluated in a Canadian cohort that includes 507 patients with CD (Jewish, n = 135; non-Jewish, n = 372), 216 patients with UC, and 352 ethnically matched controls (Jewish, n = 88; non-Jewish, n = 264) were genotyped for *SLC22A4* C1672T (rs1050152), *SLC22A5* G-207C (rs2631367) and the major CD-associated*CARD15* variants (Newman et al., 2005).

The diagnosis of CD or UC and determination of disease location were based on at least of 2 clinical episodes with confirmatory endoscopic, radiologic or histologic features (Newman et al., 2005). Sites of disease were determined by colonoscopy, barium studies and histologic findings, with ileal involvement based on small bowel enema and/or laparotomy findings (Newman et al., 2005). Since many patients had multiple sites of disease association, site of disease was not solely assigned and disease location was expressed as proportion of patients with any involvement at specific sites (Newman et al., 2005).

This study reported that their data showed no effect of the 2-allele risk haplotype on age at disease diagnosis. The *SLC22A*-TC haplotype was strongly associated ($P < .0001$) with CD in the non-Jewish subgroup, but not in the Ashkenazi Jewish white population (Newman et al., 2005). The combination of *SLC22A*-TC homozygosity and one or more of the common *CARD15* disease susceptibility alleles produce a 7.5-fold increase in risk for CD ($P = 9 \times 10^{-8}$) and a 4.5-fold increase in risk for ileal disease ($P = 0.001$). The risk haplotype showed only a suggestive association with CD in the Jewish subgroup and no association with UC in the cohort or in subgroups stratified by *CARD15* genotypes (Newman et al., 2005).

In another study, a total of 12 SNPs in the exonic sequences of the OCTN1 gene was examined and two nonsynonymous SNPs, rs1050152 (L503F) and rs272879 (L395V) were

studied with IBD in a cohort of familial and sporadic IBD from the central Pennsylvania, USA area (Lin et al., 2010). The study samples (n=465) included 212 inflammatory bowel disease patients (CD=115, UC=97), including 103 familial (CD=55, UC=46) and 111 sporadic (CD=60, UC=51) IBD, 139 non-IBD family members from a familial IBD registry and 114 unrelated healthy controls (Lin et al., 2010). The study reported that the OCTN1 rs1050152 is associated with CD (p=0.042) and with IBD (p=0.030); whereas the new variant of OCTN1, rs272879 is not associated with IBD, CD or UC. The distribution of the rs1050152 variant showed a high level of the T allele in male UC (p=0.023) and IBD (p=0.042) patients and in female CD (p==0.039) patients (Lin et al., 2010).

Table 1: *SLC22A4* and *SLC22A5* SNPs associated with inflammatory bowel diseases.

Study	**SNPS**	**Genotypes elevating the risk**	**Association**
Peltekova et al., 2004,	*SLC22A4*, rs1050152 *SLC22A5*, rs2631367	C→T transition (C1672T) in SLC22A4. G→C transversion (G-207C) in SLC22A5.	Associated with CD.
Yamazaki et al., 2004	*SLC22A4*, rs1050152 *SLC22A5*, rs2631367	C→T transition (C1672T) in SLC22A4. G→C transversion (G-207C) in SLC22A5.	Not associated with CD.
Newman et al., 2005	*SLC22A4*, rs1050152 *SLC22A5*, rs2631367	C→T transition (C1672T) in SLC22A4. G→C transversion (G-207C) in SLC22A5.	Associated with CD but not UC.
Waller et al., 2006	*SLC22A4*, rs1050152 *SLC22A5*, rs2631367	C→T transition (C1672T) in SLC22A4. G→C transversion (G-207C) in SLC22A5.	Associated with UC.
Lin et al., 2010	*SLC22A4*, rs1050152 & rs272879	C→T transition (C1672T) in SLC22A4.	rs1050152 is associated with CD; rs272879 is not associated with CD or UC.
Martini et al., 2012	*SLC22A4*, rs1050152	C→T transition (C1672T) in SLC22A4.	Associated with CD.

To date, many studies have been reported in the literature determining the association of *SLC22A4* and *SLC22A5* with CD, reporting the same SNP rs1050152 and rs2631367 (Armuzzi et al., 2003; Newman, 2005; Noble et al., 2005; Peltekova et al., 2004; Török, H-P et al., 2005; Vermeire et al., 2005).

Among these studies few of them could not find any associations, however majority of these studies have seen evidence of association reinforcing the fact *SLC22A4* and *SLC22A5* are associated with CD (Table 1) (Noble et al., 2005; Vermeire et al., 2005; Yamazaki et al., 2004).

Conversely, very little research has been done with *SLC22A4/ SLC22A5* and UC and most of them found no association with UC (Newman et al., 2005; Torok et al., 2005). For the first time in a well powered study an association between OCTN variants and UC was demonstrated by Waller et al. in 2006. Due to various conflicting reports of the association between markers in the *IBD5* region and UC the aim of Waller et al. study was to ascertain the contribution of OCTN variants to UC and CD in a large independent UK dataset. The study also tried to identify interactions between the *IBD5* and *CARD15* loci and to seek genetic evidence that the OCTN association is distinct from the IBD5 risk haplotype (Waller et al., 2006).

A total of 1104 unrelated Caucasian IBD patients of north European origin, that includes 496 CD, 512 UC, 96 indeterminate and 750 ethnically matched controls were genotyped for three single nucleotide polymorphisms (SNPs) in the CD associated genes OCTN1+1672 (rs1050152), OCTN 22207 (rs2631367) and IGR2230 (rs17622208) and two flanking IBD5 tagging SNPs, IGR2096 (rs12521868) and IGR3096 (rs7705189) (Waller et al., 2006).

The study reported that OCTN variants were strongly associated with both UC and CD. There was no model of interaction between *CARD15* and *IBD5*. Overall, the study concluded

that the OCTN variants were associated with susceptibility to IBD overall, as the effect was equally strong in UC and CD (Waller et al., 2006).

In another research, Martini et al. in 2012 tried to evaluate if single nucleotide polymorphisms of the *SLC22A4* OCTN1) and *SLC22A5* (OCTN2) genes are associated with ulcerative colitis (UC), sporadic colorectal cancer (SCC), and with UC cases with cancer progression (UCCP) particularly in younger patients (<55 years). The OCTN1 and OCTN2 polymorphisms were evaluated in 200 patients with UC, 59 patients with UCCP, 200 patients with SCC and in 200 controls. IL-8 expression was also assessed by real-time polymerase chain reaction (PCR) (Martini et al., 2012). In addition, human colon carcinoma Caco-2 cells were transfected with the OCTN1/1672C allele, which is homozygous for OCTN1/1672T variant. NF-κB activity was evaluated by luciferase based reporter assay and IL-8 mRNA expression by real-time PCR (Martini et al., 2012).

The result of the study showed that OCTN2 polymorphisms did not present a significant association with any group of patients compared to normal controls. In contrast, homozygosity for the OCTN1/1672T variant was significantly associated with UC ($P = 0.047$) with UCCP ($P < 0.001$), and with SCC developing in early age (<55 years) ($P = 0.021$). Importantly, IL-8 mRNA expression was higher in UC and UCCP patients homozygous for the OCTN1/1672T variant compared to the other genotypes. Furthermore, in Caco-2 cells transfection of the OCTN1/ 1672C variant reduced the activity of the proinflammatory factor NF-κB (Martini et al., 2012). Overall, the study demonstrated that OCTN1/1672T variant is associated with an increased risk of UC (Martini et al., 2012).

Secondly, *SLC22A5* (solute carrier family 22, member 5) is a membrane transporter protein that was first cloned in 1998 in parallel from rat and man (Sekine et al., 1998; Tamai et al., 1998; Wu et al., 1998). It is located in chromosome 5q23.3 with 25,906 bases length. The

protein encoded by *SLC22A5* is known as OCTN2 (Wu et al., 1998). In 1998, expression of *SLC22A5* in HEK293 cells showed high transport for carnitine in a sodium dependent manner (Tamai et al., 1998).

Being a Na+/carnitine cotransporter with a high affinity for carnitine, human OCTN2 operates as polyspecific, Na+ independent organic cation transporter that can transport cations in both directions across the plasma membrane (Koepsell et al., 2007).

Human OCTN2 also translocates acetyl-L-carnitine, buturyl-Lcarnitine (Srinivast al., 2007), pivaloylcarnitine, valproylcarnitine (Ohnishi et al., 2008), the cardioprotective drug mildronate (Grigat et al., 2009), the b-lactam antibiotic cephaloridine and emetine in a sodium dependent manner (Koepsell et al., 2007). Human OCTN2 protein has been localized to the brush-border membrane of renal proximal tubule cells and small intestinal enterocytes (Koepsell et al., 2007).

The gene is expressed in a wide range of tissues. It was strongly expressed in kidney, skeletal muscle, placenta, heart, prostate, and thyroid and weakly in pancreas, liver, lung, brain, small intestine, uterus, thymus, adrenal gland, trachea, spinal cord, and several other tissues (Tamai et al., 1998). The amino acid sequence of human OCTN2 is 75.8% similar to that of human OCTN1 (Tamai et al., 1998).

Genetic variations in the *SLC22A5* gene causes reduced expression or impaired function of OCTN2 that leads to a recessively inherited disorder called “primary systemic carnitine deficiency” (SCD) (Koepsell et al., 2007). Due to the role of carnitine in the gastrointestinal tract became an important topic of discussion, as several genetic studies linked mutations in genes coding for carnitine transporters OCTN1 (*SLC22A4*) and OCTN2 (*SLC22A5*) with CD (Shekhawat et al., 2007).

In 2004 Peltekova et al. study, Patients with CD have been shown to have a missense substitution 1672C→T in OCTN1 causes amino acid substitution L503F, and a G→C transversion in the promoter region of OCTN2 (-207 G→C), which disrupts a heat shock binding element (HSE). Both of these mutations are in strong linkage disequilibrium and generate a two-allele risk haplotype. These mutations showed a reduced carnitine uptake by OCTN1 and a reduced expression of OCTN2 in *in-vitro* experiments, thus potentially causing tissue carnitine deficiency (Peltekova et al., 2004). This association study suggests that carnitine transport deficiency might play an important role in the pathogenesis of Crohn's disease (Shekhawat et al., 2007).

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 principally important for the enterocytes to maintain normal gut morphology and function. The role of carnitine in the pathogenesis of intestinal inflammation is supported by Shekhawat et al. in 2007. The study included a mouse model with functional defect in OCTN2 carnitine transporter (Shekhawat et al., 2007). The levels of carnitine in the gut were reduced by 90% in this $Slc22a5^{-/-}$ knockout mouse model which resulted in intestinal villous atrophy and intestinal breakdown and inflammation with intense lymphocyte and macrophage infiltration, which in turn lead to ulcer formation and apoptosis of gut epithelial cells (Shekhawat et al., 2007).

This study showed that carnitine has an obligatory role in the maintenance of normal intestinal and colonic structure and morphology and also carnitine-dependent oxidation of long-chain fatty acids in mitochondria is essential for normal gut function. The study suggested that in order to boost fatty acid oxidation, carnitine supplementation may be therapeutically beneficial in intestinal inflammation patients (Shekhawat et al., 2007).

In 2009, Wojtal et al. determined the mRNA expression levels of 15 solute carrier transporters (including OCTN1 and OCTN2) in two regions of the intestine in 107 IBD patients (53 with Crohn's disease and 54 with ulcerative colitis) and 23 control subjects (Wojtal et al., 2009).Endoscopic biopsy specimens were taken from two locations (terminal ileum and colon) for RNA extraction and histological examination. mRNA expression was evaluated using the quantitative reverse transcription-polymerase chain reaction technique (Wojtal et al., 2009). The result of the study showed that mRNA levels for organic zwitterion/cation transporter, OCTN2, were significantly lower in both ileum and colon of IBD patients. The study concluded that in IBD patients intestinal SLC mRNA levels are dysregulated, which may be correlated to the inflammation of the tissue and provides an indication about the role of inflammatory signaling in regulation of SLC expression (Wojtal et al., 2009).

In 2008, Noble et al. showed that the mRNA expression levels of OCTN1 and OCTN2 were significantly decreased in patients with UC (Noble et al., 2008).

Although various studies showed the association of *SLC22A4* and *SLC22A5* in the development of IBD; the actual mechanism by which these genes triggers intestinal inflammation remains unclear.

However; the above mentioned association studies suggests that carnitine transport deficiency in *SLC22A4* and *SLC22A5* might play a role in the pathogenesis of IBD (Shekhawat et al., 2007). For instance, Peltekova et al. study showed that mutations in *SLC22A4* and *SLC22A5* cause a reduced carnitine uptake potentially causing tissue carnitine deficiency (Peltekova et al., 2004). Shekhawat et al showed carnitine has an obligatory role in the human body and supplementation may be therapeutically beneficial in intestinal inflammation patients (Shekhawat et al., 2007). And in a randomised clinical trial, Mikhailova et al. showed that a 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 (Mikhailova, et al., 2011).

In contrast, when ergothioneine is considered as the main substrate for SLC22A4, in 2005, Taubert et al., suggested that carriers of the rs1050152-T allele accumulate higher ergothioneine concentrations in OCTN1 expressing cells compared to the wild-type rs1050152-C allele. Therefore, high concentration of ergothioneine is involved in the CD development (Taubert et al., 2005). The same study also tested the effect of ET on proliferation of the Caco-2 cell line. The results showed that rather than antioxidant activities, stimulatory effects on cell proliferation appear to constitute the functional role of ergothioneine. And ergothioneine may accelerate the inflammatory process by transcriptional activation of fibroblast repair proliferation, thus also conferring susceptibility of CD patients to develop colorectal cancer (Taubert et al., 2005).

In agreement with this theory, in 2009 Taubert et al. performed another study analyzing whether ergothioneine concentration in CD affected tissues are related to the genotype and expression of *SLC22A4* (Taubert et al., 2009). And the result showed that in inflamed mucosal biopsies from CD patients, ergothioneine and *SLC22A4* levels were elevated by two fold in carriers of the rs1050152-TT genotype compared with normal mucosa (Taubert et al., 2009). High ileal expression of *SLC22A4* and the inflammation-triggered upregulation of *SLC22A4* or ergothioneine were also observed (Taubert et al., 2009). Overall, the study proposed a possible functional role for ergothioneine in the pathophysiology of CD. However; restricted presence of ergothioneine in wide variety of foods may contradict this theory.

While substrate impairment is clearly a factor but location of mutations in the *SLC22A4* and *SLC22A5* within the *IBD 5* locus (IBD5) is also assumed to have significant contribution in

the mechanism of IBD (Peltekova et al., 2004). Inflammatory bowel disease 5 or *IBD5* (OMIM ID 606348) is a highly conserved 250 kilobase haplotype of 11 SNPs spanning a cytokine gene cluster; located on chromosome 5q31 that is associated with an increased risk of CD (Rioux et al., 2001). It was originally identified in 1999 as conferring susceptibility for CD (Ma et al., 1999; Noble et al., 2005). However linkage for CD on 5q31 was first reported in the Canadian population by Rioux et al. in 2000.It is contrary, if this locus is also associated with UC as only two studies have seen association (Giallourakis et al., 2003 & Noble et al., 2008) and majority of the other research studies have not seen any evidence of association with UC (Armuzzi et al., 2003; Mirza et al., 2003; Negoro et al., 2003).

The *IBD 5* region contains a number of potentially genes relevant to IBD development (Figure 2), including several genes encoding for immunoregulatory cytokines (Weersma et al., 2007) and only two membrane transporter genes *SLC22A4* and *SLC22A5* (Peltekova et al., 2004; Rioux et al., 2000). It is acknowledged that SNPs at the IBD5 locus are highly associated with Crohn's disease, including *SLC22A4* and *SLC22A5* (Fisher et al., 2006; Repnik and Potocnik, 2011).

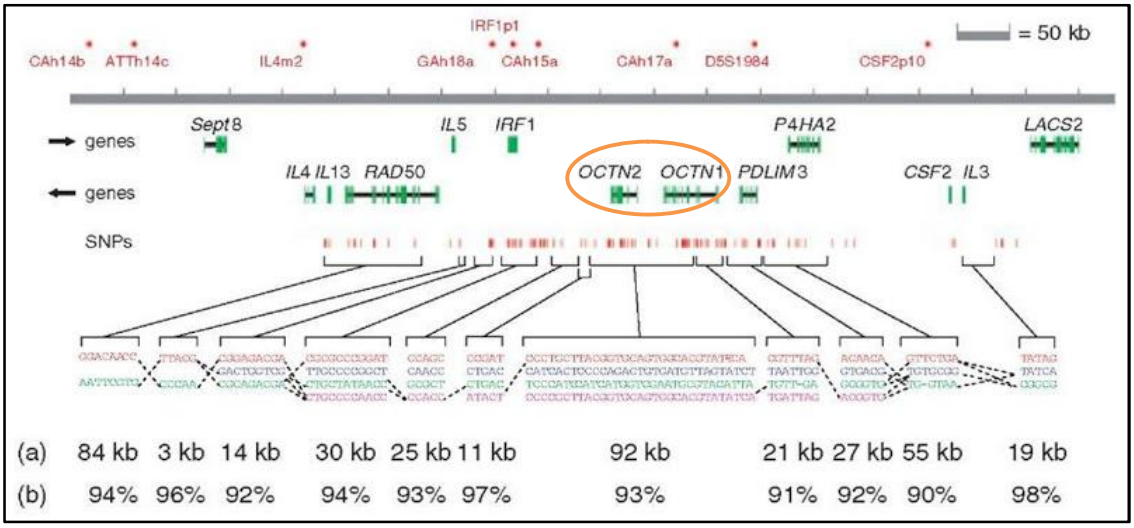


Figure 2: Haplotype structure at the IBD5 locus (Daly et al., 2001). OCTN1 and OCTN2 highlighted with orange circle.

It is assumed genetic variation on the risk haplotype could influence the function or expression of one or more genes in the *IBD5* region (Reinhard and Rioux, 2006). However, the proper mechanisms through which genetic variation in IBD5 alters physiological processes are unknown (Repnik and Potocnik, 2011).

Besides *SLC22A4* and *SLC22A5,* the *SLC22A23* gene has been associated to IBD (Barrett et al., 2008; Franke et al., 2008; Palmieri et al., 2011). In latest work, one of our lab member was able to show that homozygotes for SNPs rs4959235-TT and rs950318-GG were exclusively found in individuals with UC, resulting in odds ratios of 11.42 (95% CI: 0.63, 266; p=0.03) and 16.71 (95% CI: 0.95, 294; p=≤0.001) respectively (Serrano et al., 2014).

Based on the existing evidence, we hypothesize, that imbalances in organic cations, either of endogenous or dietary origin, contribute to IBD susceptibility (Pochini et al., 2013) and that dysregulations of *SLC22A4* or *SLC22A5* are disease causative.

In order to validate genetic association studies functional studies need to confirm an impact of the genetic variation on the genes or proteins function. In the case of *SLC22A23* this is at present not being done, even its basic biology remains to be demined.

4.4. Solute Carrier Family 22, Member 23 (*SLC22A23*)

The proposed research focuses on a gene called *SLC22A23*, a relatively uncharacterized member of solute carrier family 22, were first identified by Jacobsson et al. in 2007. The human orphan transporter *SLC22A23* was identified through data base analysis and was classified within the phylogenetic tree of organic cation transporters. It is located in human chromosome 6p25.2 and has 7 exons (Jacobsson et al., 2007). Very little is known about its physiological function because the genomic organization, localization and genetic variations

have not been described. However, Bennett et al. investigated the rat *slc22a23* ortholog but the substrate for this transporter is yet to be determined (Bennett et al., 2011).

Although the gene's physiological function is poorly characterized some common and complex diseases have been associated with genetic variations in the *SLC22A23* locus, including recently inflammatory bowel disease (IBD).

4.5. *SLC22A23* Gene And Its Orthologues In Murine Species

There is not much known about *SLC22A23* because its physiological function, structure, transport characteristics and functional studies have not been previously reported.

The human orphan transporter *SLC22A23* was first described by Jacobsson et al; in 2007. It was identified through data base analysis and classified within the phylogenetic tree of organic cation transporters (Jacobsson et al., 2007).It is located in human chromosome 6p25.2, has 7 exons and putatively encodes a protein of 405 amino acids (Jacobsson et al., 2007). In mouse it is located in 19qA3.3, has 9 exons, putatively encoding a protein of 689 amino acids (Jacobsson et al., 2007).In rat it is located in 17p12, has 9 exons and putatively encodes a protein of 689 amino acids (Jacobsson et al., 2007). A tissue expression study for *slc22a23*was performed on a panel of 30 rat tissues using quantitative real-time polymerase chain reaction. The results showed a broad tissue distribution in spinal cord, kidney, liver, eye, adipose tissue, lung, epididymis, adrenal gland, pineal gland, skeletal muscle, heart, spleen, thymus, ovary, uterus, testis, and epididymis with the exception of intestine and with highest expression level in liver (Jacobsson et al., 2007).

The study also reported that *slc22a23* lack close paralogs and are phylogenetically isolated from the rest of the family. This is because unlike other *SLC 22* members this gene consists of 9 transmembrane (TM) regions and lack the first 3 transmembrane (TM) and the large

extracellular loop. Hence it is considered that these proteins may have alternative transport mechanisms and binding sites. The study also confirmed that rodent *slc22a23* is orthologues to human *SLC22A23* (Jacobsson et al., 2007).

The second study was published in 2011 where rat *slc22a23* was cloned and characterized Bennett et al., 2011). Northern blot analysis showed that *slc22a23* mRNA was expressed in a wide variety of tissues. The highest levels were observed in liver, brain, kidney, and spleen. Lower levels were observed in skin, lung, and intestine. Expression was low in the remaining tissues (Bennett et al., 2011). *Slc22a233* cDNA was isolated from rat brain capillary and from cultured neurons using PCR techniques. For functional analysis, the cDNAs of the gene was subcloned into a plasmid expression vector and transfected into HEK-293 cells, as were control cDNAs. Recombinant cell surface protein was verified by western blot and fluorescence microscopy. Recombinant protein was found on the cell membrane and in the cytoplasm (Bennett et al., 2011).

Sequence analysis of *slc22a23* and comparison with other organic cation transporter was done. The sequences found for *slc22a23* was more similar to *slc22a17* (34% identity) than to other *slc22* genes. The greatest difference between the gene and the rest of the SLC22 family is at the N-terminus. After performing 5' and 3' RACE (rapid amplification of cDNA ends), nearly full-length mRNA sequence for *slc22a23* was determined showing the gene mRNA contains an unusual GC-rich region at the 5' end. *Slc22a23* was found significantly larger (681aa) than the typical slc22 protein (Bennett et al., 2011).

Transport activity of the gene was evaluated using radioisotope uptake assays. The range of compounds tested included all known SLC22 substrates.

Despite sequence homology to the OCT family, slc22a23 did not transport typical organic cations like 1-methyl-4-phenylpyridinium (MPP+), Tetraethylammonium (TEA) and choline,

anion transporters like para-aminohippurate (PAH) and ibuprofen, zwitterions like arginine, nor fluorescent cations like ethidium, hence showing a more restricted substrate profile than other polyspecific members (Bennett et al., 2011).

Several other substrates were tested for transport activity such as neuro-transmitter transport, urate, nicotinamide and ascorbic acid. To test if the lack of functional activity was a consequence of using HEK293 cells the gene was expressed in other cell lines as well such as 3T3, MC-IX-C or MDCK cells however no transport activity was observed (Bennett et al., 2011).

Thus, *slc22a23* was unable to transport typical substrates for the slc22 family showing that the gene may have restricted substrate profile than other polyspecific members. It was also speculated that the gene may require some type of cofactor to show functional activity (Bennett et al., 2011).

4.6. Disease Associations Of *SLC22A23*

Till date some diseases has been associated with *SLC22A23*genetic variations, such as cardiovascular diseases, breast neoplasms and inflammatory bowel disease (IBD).

A study, published in 2010, detected the genetic variation mediating the antipsychotic-induced individual variation in QTc prolongation in 738 schizophrenia patients (Aberg et al., 2012). The outcome was brilliant in achieving genome-wide significance, defined as a Q-value ≤0.10 (P-value¼1.5410, Q-value¼0.07) with SNP rs4959235 in*SLC22A23* that mediated the effects of quetiapine on QTc prolongation (Aberg et al., 2012). The study also reported that as an organic ion transporter, the mechanism of *SLC22A23*'s involvement in antipsychotic-induced QTc prolongation could be by means of the clearance of the drug from the heart, or by shuttling of molecules involved in cardiac function (Aberg et al., 2012).

SLC22A23 has been linked to breast neoplasm in a study using a Cox proportional hazards model to predict recurrence of triple negative breast cancer in Taiwanese patients (Chen, 2011). In this study, six prognostic genes expression profile were identified along with *SLC22A23* for predicting the risk of recurrence of triple-negative breast cancer (TNBC) in Taiwanese patients (Chen, 2011).

Zhao et al. (2011) showed associations between Wolf–Hirschhorn syndrome candidate gene1 (WHSC1) alleles and endometriosis-related infertility and *SLC22A23* haplotypes associated to the severity of disease stages (Zhao et al., 2011). An association between *SLC22A23* haplotypes with SNPs in miR-125a-3p binding sites and pre-disposition to endometriosis was found (Zhao et al., 2011).

The *SLC22A23* target site harbors five miRNA target SNPs: rs3813486, rs1127473, rs7742745, rs35091219 andrs3211066. Mutation in the functional SNP rs7742745 was only detected in three cases and not in controls, suggesting a possible role in risk for endometriosis in patients. Further haplotype analysis for three SNPs (rs3813486, rs1127473 and rs3211066), excluding the rare variants, supported a role for variation in *SLC22A23* in endometriosis risk. Haplotypes in *SLC22A23* were strongly associated with infertility and more severe stage of the disease (Zhao et al., 2011).

Like *SLC22A4* and *SLC22A5* (Pochini et al., 2013), *SLC22A23* has also been associated to inflammatory bowel diseases (IBD) few times(Barrett et al., 2008; Franke et al., 2008;McGovern et al., 2010; Palmieri et al., 2011; Serrano et al., 2014). While few of these studies are genome-wide association studies (GWAS) (Barrett et al., 2008; Palmieri et al., 2011). GWAS is type of study where one compares the allele frequency of a particular variant between unrelated cases and controls (Ek Weronica et al., 2014). That is, it is an examination of many common genetic variants in different individuals to see if any variant is

associated with a trait. It specifically focuses on associations between single-nucleotide polymorphisms (SNPs) and traits similar to major diseases. These studies normally compare the DNA of two groups of participants: people with the disease (cases) and similar people without the disease (controls). Each person gives a sample of DNA, from which millions of genetic variants are read using SNP arrays. If one type of the variant (one allele) is more frequent in people with the disease, the SNP is said to be "associated" with the disease. The associated SNPs are then considered to mark a region of the human genome which influences the risk of disease. In contrast to other methods the GWA studies investigate the entire genome (Manolio, 2010; Pearson and Manolio, 2008).

4.7. Associations Of *SLC22A23* To IBD

A meta-analysis of genome-wide association studies, published in 2011, investigated the level of transcript expression of the previously 71 depicted loci in the colonic mucosa biopsies of patients with CD, during an active phase of disease (Palmieri et al., 2011). Sixty-seven deregulated genes out of 71 loci were identified analyzing the gene expression in the colonic mucosa of CD. Among 67 genes, 16 are new additional genes including *SLC22A23* that was reported to be associated with active phase CD. Extensive resequencing, large-scale fine mapping and functional studies of these loci will be required to elucidate the pathogenic mechanisms of CD (Palmieri et al., 2011).

In another study published in 2008, a meta-analysis of three separate genome-wide association scans for CD in European-derived populations was performed that reported SNP rs17309827 in *SLC22A23* is nominally associated with CD (Barrett et al., 2008).

The same SNP (s17309827) in *SLC22A23* was also reported to be significantly associated (P=0.05 after Bonferroni correction for multiple testing) with ulcerative colitis or Crohn's

disease (Franke et al., 2008). The study included German sample of individuals with Crohn's disease (n=1850) or ulcerative colitis (n=1103) and healthy controls (n=1817) (Franke et al., 2008).

SLC22A23 was also associated with UC, a chronic, relapsing inflammatory condition of the gastrointestinal tract with a complex genetic and environmental etiology (McGovern et al., 2010). CD specific SNP rs17309827 was associated with UC (McGovern et al., 2010) in the ethnically distinct north Indian population (Juyalet al., 2011). In contrast, a study analyzing 45 single nucleotide polymorphisms, tagging 29 of the loci recently associated with CD in 2527 UC cases and 4070 population controls found no association of SNP rs17309827 with UC (Anderson et al., 2009). For some loci, this may reflect a lack of power because extremely large sample sets would be required to detect small effects of UC (Anderson et al., 2009).

Recently a research in our lab performed by one of our lab member determined the association of some common and rare single nucleotide polymorphisms (SNPs) in *SLC22A23* to IBD risk in a Manitoba cohort. Selected genetic variations were genotyped in 160 CD, 149 UC, and 142 healthy controls. Statistical analysis, genetic associations, linkage disequilibrium (LD) and haplotype analysis was performed. The study for the first time reported strong associations of the two genotypes rs4959235-TT and rs950318-GG in individuals with UC (Serrano et al., 2014).

There is a strong evidence *SLC22A4* and *SLC22A5* has been associated to IBD multiple times (Franke et al., 2008; Barrett et al., 2008 & Palmieri et al., 2011). But recently a new gene *SLC22A23* is associated to IBD (Serrano et al., 2014) which needs to be further investigated since its genetics, functional studies and genomic localization are still undetermined.

4.8. Knowledge Gap

Although associated to IBD, the human *SLC22A23* is an orphan gene encoding a putative membrane transporter protein.

The genomic architecture of the human *SLC22A23* gene is not described, nor is the expression pattern, differentially spliced transcripts, and genetic variations. Therefore, protein isoform, as well as intracellular targeting are not characterized. The substrate remains to be determined.

4.9. Rational

To validate the contribution of *SLC22A23* for IBD it is significant to characterize the gene, determine its cellular localization and most importantly find out the substrate to understand the molecular mechanism of *SLC22A23´*s contribution to IBD.

4.10. Hypothesis

The *SLC22A23* gene encodes a membrane transporter relevant in the aetiology of inflammatory bowel disease.

Sub-Hypothesis:

- The *SLC22A23* gene has a distinct chromosomal location, genomic architecture and expression pattern.
- The SLC22A23 protein isoforms locate to the plasma membrane.
- The SLC22A23 substrate is a metabolite, defined as a small solute molecule.

5. RESEARCH PLAN

5.1. Objectives

The proposed research addresses the knowledge gaps and aims to characterize the following:

1. To characterize the genomic structure of the human *SLC22A23* gene.
2. To determine the tissue expression of the *SLC22A23* in human.
3. To determine the location of the protein isoform in mammalian cells CHO, MDCK and CaCo- 2.
4. To establish stable cell lines expressing the protein isoform.
5. To determine the substrate of the SLC22A23 protein using aLiquid Chromatography-Quadrupole Time of Flight-Mass Spectrometry (LC-QTOF-MS), if possible.

6. EXPERIMENTAL METHODS

To obtain the above objectives the following method were used:

6.1. Genomic Characterization Of The Human *SLC22A23* Gene Using Databases

The genomic characterization of *SLC22A23* gene was performed using databases. Information about the *SLC22A23* gene is contained in the National Center for Biotechnology Informations (NCBI), Ensembl, Ace View and UCSC databases. Under NCBI and Ensembl we narrowed the search under "Gene", using the key words *SLC22A23* and under the sub heading - the chromosomal location, chromosomal neighbourhood, length, and number of exons of the gene. We have compared the information from different databases.

6.1.1. Chromosomal Localization And Gene Neighborhood - The chromosomal localization and the gene neighborhood of *SLC22A23* was searched in Gene Card, NCBI, Ensembl and UCSC genome browser.

6.1.2. Gene Structure - The gene length, number of exons and intron structure of *SLC22A23* was searched in NCBI, Ensembl and UCSC genome browser (April, 2015).

6.1.3. Splice Variants - Number of Splice variants of the *SLC22A23* gene was searched in the Ensembl databases.

6.1.4. Genetic Variations - Genetic variations of the *SLC22A23* gene were searched in the Ensembl databases under variation table we searched for high impact variations only such as splice donor variant, splice acceptor variant, stop gained frameshift variant and missense single nucleotide polymorphisms (SNPs). We calculated the repeated number of transcripts and recreated a new table variation.

6.1.5. Tissue Expression - Tissue expression of the *SLC22A23* gene was determined using UniGene EST (Expressed sequence tags) profile for cluster Hs. 713588 (http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Hs.713588) and also from Ace View (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&term=slc22a23&submit=Go).Tissue expression data has been represented by a bar chart using Microsoft Excel with tissues against Transcripts per million.

6. 2. Genomic Characterization Of The Human *SLC22A23* Gene Using Bioinformatics Analysis

Human genomic sequence and the reference RNAs for *SLC22A23* gene was downloaded from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/gene/63027) in FASTA files from human RefSeqs 37.3.

The Basic Local Alignment Search Tool (BLAST) and the Ensembl data base (http://uswest.ensembl.org/ ENSG00000137266) was used to screen for human non-reference RNAs as well as expressed sequence tags (ESTs). To increase the accuracy of the search, the reference RNA sequence was broken into shorter sequences. The DNA Sequence Analysis Software - Sequencher 5.0 (Gene Codes Corporation; Ann Arbor, USA) was used to align and hand curate the remaining sequences. Finally, the resultants were compared to the consensus sequences and the anomalies analyzed. Exons were deduced from the alignments of all transcripts (RNAs and ESTs) in order of appearance on the RNA sense orientation in the *SLC22A23* locus.

6. 3. Tissue Expression Of The Human SLC22A23 Using Real-Time Polymerase Chain Reaction

Tissue expression of the human *SLC22A23* was performed using Real-Time polymerase chain reaction, on a Step One Plus™ (Applied Biosystems) real time PCR thermocycler. Brain, lung, pancreas, kidneys, small and large intestine, as well as the intestinal segments esophagus, stomach, liver, colon ascending, colon transverse, colon descending, cecum, illeocecum, Illeum, jejunum, duodenum and rectum cDNAs was used in the experiment (Table 2).

Table 2: Identity of cDNAs used in the tissue expression.

Name of first strand cDNA	Reference No (Clontech)	Lot No
Liver	S1691	3030144
Kidney	637204	7030190
Colon Transverse	S1683	3030136
Colon Ascending	S1681	3030134
Colon Descending	S1682	3030135
Cecum	S1689	3030142
Illeocecum	S1685	3030138
Illeum	S1688	3030139
Jejunum	S1686	3030140
Duodenum	S1684	3030137
Stomach	S1690	3030143
Esophagus	S1692	3030145
Rectum	S1687	3030141

Assays were performed in triplicate in the 96 well format using TaqMan® GTXpress™ Master Mix Assay (4401892) following the manufacturer's protocol (ABI Taqman assay). 20ul volume of each samples were run at thermal cycle conditions of: 95ºC for 20 seconds for

initial denaturation followed forty cycles of 95 ºC for 3 seconds and for 50ºC for 30 second (Table 3). Data were processed using Step One™ Software v2.2.2. A Standard curve was generated from the plasmid GFP-SLC22A23-pcDNA™3.2/V5-DESTpcDNA in concentrations from 100ng/µl to 1 fg/µl. A bar chart was created using Microsoft Excel software with number of copies of cDNA per µl against the tissues.

Table 3: Thermal cycle conditions for RT-PCR.

Instrument	Step	Temp	Duration	Cycles
StepOne™ StepOnePlus™ 7900	Enzyme Activation	95ºC	20 sec	Hold
	Denature	95ºC	3 sec	40
	Anneal/Extend	60ºC	30 sec	

6.4. Cellular Localization Of The Human SLC22A23 Using PCR and Gateway® Cloning Technology

6.4.1. Sub-Cloning – *SLC22A23* transcript (NM_021945.6) was previously sub cloned in our laboratory, where *SLC22A23*open reading frame was primarily fused to a N-terminal fluorescent proteins using PCR. In this case *SLC22A23* transcript was tagged to the green fluorescent protein tGFP. The terminal ends of the amplicons were tagged with the Gateway® recombination sites AttB1 and AttB2. Then tGFP tagged *SLC22A23* gene was integrated into the donor vector pDONR™ 221 and transferred into the expression vector pcDNA^{TM}3.2/V5-DEST by Gateway LR cloning reaction following the manufacturer protocol (Figure 3).

The identity of the insert was verified by sequencing at The Centre for Applied Genomics (TCAG) in Toronto. All these steps were previously done in our laboratory.

To obtain transgenic cell lines with stable integration of the expression plasmid, tGFP-SLC22A23-pcDNA™3.2/V5-DEST was linearized to increase the chances that the vector does not integrate in a way that disrupts elements necessary for expression in mammalian cells. To linearize the construct, the vector was cut at a unique site that is not located within gene of interest using Apa I enzyme.

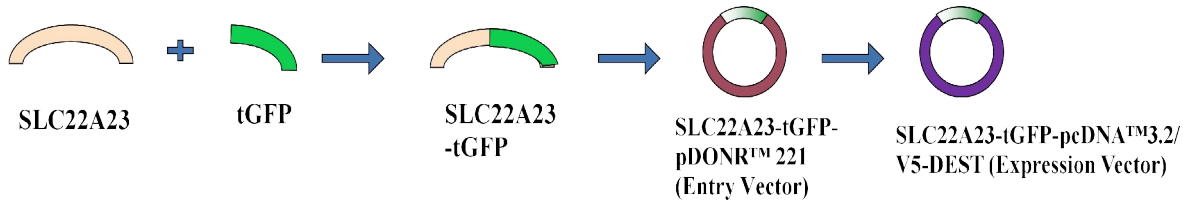


Figure 3: Gateway® Cloning Technology

6.4.2. Transfection – To verify human *SLC22A23* is a transmembrane transporter protein the tGFP-SLC22A23-pcDNA™3.2/V5-DEST transiently expressed in three mammalian cell lines CHO, MDCK and CaCo-2 using effectine transfection reagent (Invitrogen), lipofectamine transfection reagent (Invitrogen), TransIT®-LT1Transfection Reagent (Mirus Bio LLC) or TransIT-X2® Dynamic Delivery System (Mirus Bio LLC). Chinese hamster ovary (CHO) cells are the most widely used mammalian cell line. (Butler and Meneses-Acosta, 2012). Economic importance and ease of cultivation and alleged immortality make CHO cells a desirable candidate for metabolic studies (Nicolae et al., 2014). CHO cells are genetically stable, they can be kept in suspension cultures, they can be reproduced with expression vectors containing the "gene of interest" (GOI); they are easy to transfect and they remain stable during the process of selection, amplification, single-cell cloning and the characterization of the clone (Chusainow et al., 2009).

Alternatively, Madin-Darby Canine Kidney (MDCK) cells are widely used as models for epithelial cells as they have clear apico-basolateral polarity, well defined cell junctions and a rapid growth rate (Dukes et al., 2011). And CaCo-2 cell line is a continuous cell of heterogeneous human epithelial colorectal adenocarcinoma cells (Moise P et al., 1983). When cultured under specific conditions CaCo-2 becomes differentiated and polarized, and morphologically and functionally the cell resembles the enterocytes lining of small intestine (Hidalgo et al., 1989; Moise P et al., 1983). The human intestinal Caco-2 cell line has been extensively used over the last twenty years as an in vitro model of the intestinal barrier (Sambuy et al., 2005).

All three cell lines were cultured at 37^{o}C in DMEM/F-12 HyClone media (Thermo Fisher Scientific), 10% FBS (Fetal bovine serum) (Thermo Fisher Scientific) on separate 6-well tissue culture plates (Thermo Fisher Scientific, BioLite). After twenty-four hours of seeding the cells, at cell densities between 60%-80%, cells were transfected with tGFP-SLC22A23-pcDNA™3.2/V5-DEST using either Effectine or Mirus transfection reagent following the manufacturer´s protocols (QIAGEN and Mirus).

After transient transfection all three cells were incubated at 37°C and were allowed to grow in order to express the protein at least 24 hours to 48 hours. Fluorescence assay was performed 36 to 48 h following transfection using bright-field and fluorescent microscopy to see the localization of the protein.

6.4.3. Imaging – To determine cellular targeting the green fluorescence tagged *SLC22A23* Zeiss Observer Z1inverted microscopy was used with Axio Cam HRM camera at 20X objectives. Once nice live green fluorescence cells were observed; images were captured and processed with the Zen Pro 2012 software.

6.5. Generation of Monoclonal Stable Cell Lines

6.5.1. Dilution Method – The purpose of using this method is to isolate each individual cell that carries selection by plating them at very low cell densities (< 1 well per well in 96 well plates) and expand colonies from those single cells in separate wells.

In this experiment after the cells expressed fluorescence, the standard media was replaced with antibiotic containing media. The expression vector pcDNA™3.2/V5-DEST is resistance to neomycin, blasticidin and geneticin. Hence geneticin was used as antibiotic of choice which is an aminoglycoside, similar in structure to neomycin and blasticidin. 400 to 1000 µg/ml of geneticin was used for all three cell lines.G418 is labile at 37°C, therefore we changed the medium containing G418 every 7-10 days to compensate for loss of selection pressure. Cells should be grown for at least 3 weeks under selection pressure to avoid contamination with non- resistant cells (Lonza Stable Cell Guidelines).

The purpose of using antibiotic is that over time this selects cells that have stably incorporated the GFP plasmid into their genomic DNA. Cells normally differ in their susceptibility to G418, which may even vary with cell passage numbers. Cells that are cultured in serum-free media may require much lower G418 concentrations as compared to cells in media containing serum (Lonza Stable Cell Guidelines). In this case, all three cells line passage numbers were less than 40.

Outgrowth of resistant cells is normally observed after 1-2 weeks of selection, depending on cells types. However; we cultured the cells with medium containing geneticin for 7-10 days until colonies begin to appear. The whole process, that is, cell death and appearance of colonies were assayed using fluorescence microscopy. Same procedure was repeated for control wells (e.g., sample without expression plasmid).Once resistant colonies are

identified they are diluted and plated at very low cell densities (1 cell per well in 96 well plates.

6.5.2. Cloning Cylinder Method – Pyrex® cloning cylinder (Sigma-Aldrich, CLS31666) of height 6 mm and 8 mm diameter were used to isolate the fluorescent cells from the non transfected ones. Once satisfactory colonies have been located, a circle was drawn around them on the bottom of the dish using a marking pen. Growth medium was removed and discarded and the plate was washed with PBS twice. Using a sterile forceps cloning cylinder was picked up, gently pressed the flat bottom of the cylinder into the smooth silicone grease and place over a single cell or colony. Once it is placed properly over the cell, 200ul of trypsin was added to the cloning cylinder. After incubating the cells for 5 mins the cells were examined under the microscope and transferred to a suitable vessels (six well plate or 96 well plate) containing growth medium.

6.5.3. Fluorescence Activated Cell Sorting – Once we identified some resistant colonies of CHO, MDCK and CaCo-2 cells, fluorescence activated cell sorting (FACS) method was performed to separate transfected cells expressing the plasmid from the non-transfected ones.

Fluorescence activated cell sorting is a particular form of flow cytometry that allows a mixture of different cells to be sorted one by one into one or more containers (Herzenberg and Sweet 1976). The cells are sorted according to their specific light scattering and fluorescent characteristics (Herzenberg and Sweet, 1976). Using fluorescent-based segregation and counting, flow cytometry is a useful scientific instrument as it can provide fast and accurate quantitative recording of the fluorescent signals from each individual cell (Herzenberg and Sweet, 1976).

To prepare the cells for sorting, we washed the cells in 6 well plates using PBS (phosphate buffered saline) (Thermo Fisher Scientific) twice and then trypsinized the cells using TrypLE Express Enzyme (1X) (Thermo Fisher Scientific) and placed in 37 °C incubator for approximately 2 minutes. After the cells got detach from the surface we added small amount of medium containing serum (DMEM+FBS) to the cell suspension as soon as possible to inhibit further tryptic activity which may damage cells. All three cell samples were prepared in a single-cell suspension form at a concentration of 8-10 million per ml. Then 2ml of cell suspension were taken in a sterile tube and centrifuged the cells at 1200 RPM, 4˚C for 5 min and aspirated the supernatant. The cells were then washed with sorting buffer (1xPBS + 1% FBS + 2mM EDTA) three or four times. The sorting buffer was filtered through 0.2µm filter sterilize. The cells were then resuspended in the sorting buffer. A clump free single cell suspension in sorting buffer at the acceptable concentration range is required in order avoid cell adhesion. Once cells had been re-suspended, we filtered the cells through nylon mesh with the appropriate pore size (40-70µM) (BD Falcon, 5mL Tubes with 40 µM filter top cap P/N: 352235). This helps to eliminate clumps, prevent clogging the instrument and allow sorting to proceed properly (Flow Cytometry Protocols, 2013). Same procedure was followed for all three cell lines CHO, MDCK and CaCo-2 and for the controls.

For our experiment we used BD FACS Aria III sorting instrument at 100 micron nozzle. We performed monoclonal sorting into a 96-well plates containing collection medium (DMEM+25% FBS). Sorted cells were collected in the collection medium containing high amount of FBS to prevent the cells from drying out and dying and also to help cells recover from the sorting process. The cells were then placed in 37°C incubator, undisturbed for couple of weeks until cell confluency increased.

6.6. Functional Study

Extracts of cell line cultures will be analyzed using a Liquid Chromatography-Quadrupole Time of Flight-Mass Spectrometry (LC-QTOF-MS) to identify potential substrate(s) of the SLC22A23 protein.

7. RESULTS

7.1 Genomic Characterization Of The Human SLC22A23 Gene Using Database

7.1.1. Chromosomal Localization And Gene Neighborhood –

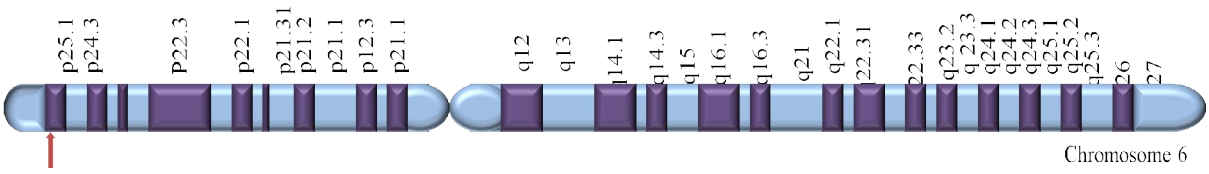


Figure 4: Chromosomal localization of *SLC22A23*. Image modified from Gene Card (April, 2015).

According to NCBI (April, 2015) and Ensembl (April, 2015), *SLC22A23,* (NM_021945.6) is located on chromosome 6, cytoband p25.2 starting at 3,268,962 base pair (bp) from the promoter region to 3,457,022 bp in the reverse strand. According to UCSC, *SLC22A23* is located on chromosome 6 with bases: 3,269,207-3,456,793. The red arrow in Figure 5 is showing the location of *SLC22A23* gene in the chromosome 6.

Table 4: Chromosome statistics. Source: Ensembl (April, 2015).

Length (bps)	**170,805,979**
Coding genes	1,043 (incl. 28 read through)
Long non coding genes	630 (incl. 5 read through)
Pseudogenes	797
Short Variants	3,885,421

The chromosome statistics shows the length of the chromosome which is 170,805,979 bps. It also shows the number of coding genes, long non coding genes, pseudogenes and short variants.

Figure 5: Location of *SLC22A23* on the chromosome 6p25.2. Source: Ensembl (April, 2015).

The red box on Figure 5 shows the location of the gene on the chromosome 6p25.2. which represents the gene is in a region of relative high gene density.

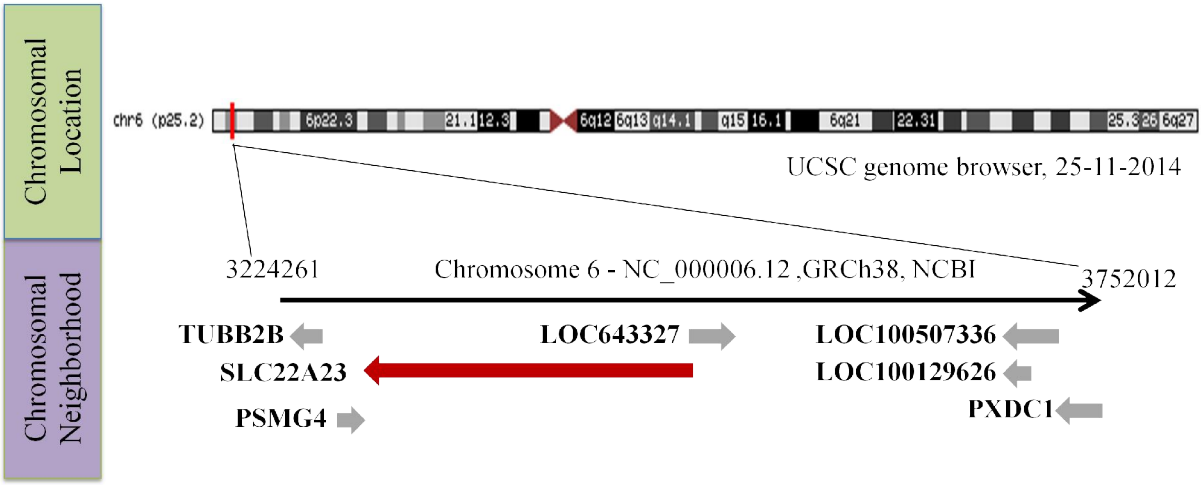


Figure 6: Gene neighborhood of *SLC22A23*. Source: Image modified from NCBI (April, 2015).

Here, *SLC22A23* is flanked by TUBB2B, tubulin beta 2B class II band, PSMG4, proteasome assembly chaperone 4 on the 5' centromere region and by LOC100507336, LOC100129626

and PXDC1, PX domain containing 1 downstream, on the 3' telomere oriented region. LOC643327 is located alongside *SLC22A23*.

7.1.2. Gene Structure –

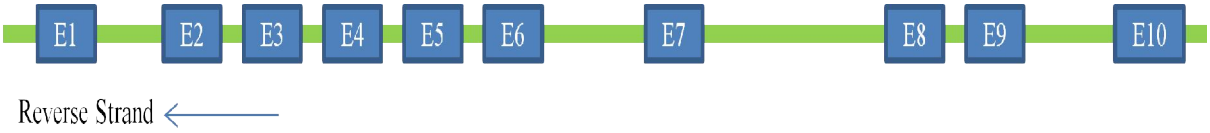


Figure 7: Exon, intron structure of *SLC22A23*. Source: Image modified from Ensembl (April, 2015).

According to UCSC genome browser the length of *SLC22A23* is 187,587 bps and encodes 11 exons. According to NCBI genome browser the length of *SLC22A23* is 187,748 bps and encodes 15 exons. According to Ace View genome browser the gene contains 23 distinct gt-ag introns, which is the splice donor and acceptor sites at the beginning of introns.

According to Ensembl genome browser *SLC22A23* has 10 exons of transcript length: 5,658 bps and translation length of 686 amino acids.

7.1.3. Splice Variants – According to Ensembl (April, 2015) genome browser, *SLC22A23* has 12 transcripts out of which 7 transcripts are protein coding, 1 nonsense mediated decay and remaining 4 are non protein isoforms (Table 5). Among the protein coding, SLC22A23-006 is of 5,658 bps in length and encodes for 686 amino acids (aa) and 10 exons. SLC22A23-009 is of 2,883 bps in length and encodes 405 amino acids (aa) and 10 exons.

Table 5: Splice variants of *SLC22A23*. Source: Ensembl (April, 2015).

Name	Transcript ID	bp	Protein	Biotype	CCDS	RefSeq	Flags
SLC22A23-006	ENST00000406686	5658	686 aa	Protein coding	CCDS47363	NM_015482 NP_056297	TSL:5GENCODE basicAPPRISCI3
SLC22A23-009	ENST00000380302	2883	405 aa	Protein coding	CCDS34331	NM_001286455 NP_001273384	TSL:1GENCODE basic
SLC22A23-005	ENST00000490273	2043	405 aa	Protein coding	CCDS34331	NM_021945 NP_068764	TSL:1GENCODE basic
SLC22A23-001	ENST00000380298	1574	361 aa	Protein coding	CCDS75389	NM_001286456 NP_001273385	TSL:1GENCODE basic
SLC22A23-002	ENST00000436008	6641	694 aa	Protein coding	-	-	TSL:5GENCODE basicAPPRISCI
SLC22A23-007	ENST00000485307	1574	440 aa	Protein coding	-	-	CDS 5' incompleteTSL:1
SLC22A23-010	ENST00000467177	1221	386 aa	Protein coding	-	-	CDS 3' incompleteTSL:1
SLC22A23-008	ENST00000497691	2555	92 aa	Nonsense mediated decay	-	-	TSL:1
SLC22A23-004	ENST00000433689	730	No protein	Processed transcript	-	-	TSL:2
SLC22A23-011	ENST00000496753	561	No protein	Processed transcript	-	-	TSL:4
SLC22A23-012	ENST00000467144	542	No protein	Processed transcript	-	-	TSL:4
SLC22A23-013	ENST00000482874	622	No protein	Retained intron	-	-	TSL:3

SLC22A23-005 is of 2,043 bps in length and encodes 405 amino acids (aa) and 11 exons. SLC22A23-001 is of 1,574 bps in length and encodes 361 amino acids (aa) and 4 exons. SLC22A23-002 is of 6,641 bps in length and encodes 694 amino acids (aa) and 11 exons. SLC22A23-007 is of 1,574 bps in length and encodes 440 amino acids (aa) and 9 exons. SLC22A23-010 is of 1,221 bps in length and encodes 386 amino acids (aa) and 10 exons.*SLC22A23*-008 is a nonsense mediated decaysplice variant which is of 2,555 bps in length and encodes 92 amino acids (aa) and 12 exons. Among the non protein coding, *SLC22A23*-004 is of 730bps in length and encodes 2 exons. *SLC22A23*-011is of 561 bps in length and encodes 4 exons. *SLC22A23*-012 is of 542 bps in length and encodes 5 exons. And the last splice variant, *SLC22A23*-013 is a retained intron biotype is of 622 bps and encodes 3 exons.

7.1.4. Genetic Variations – In total there are 71065 SNPs reported till date for *SLC22A23* according to Ensembl database (April, 2015).

Table 6: Genetic Variation of SLC22A23. Source: Ensembl (May, 2015).

No of variant consequences	Type	Description
7	Splice donor variant	A splice variant that changes the 2 base region at the 5' end of an intron
1	Splice acceptor variant	A splice variant that changes the 2 base region at the 3' end of an intron
6	Stop gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript
4	Frameshift variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three
128	Missense variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved

Here we prioritize and mentioned the most important ones which include splice donor variant, splice acceptor variant, stop gained, frameshift variant and missense variant (Table 6). Find a detailed listing of the variants in the appendix (Section 10.1, Tables a-e).

7.1.5. Tissue expression –

a) Unigene browser: Unigene (April, 2015) shows the expressed sequence tag (EST) profiles which shows approximate gene expression patterns as inferred from EST counts and the cDNA library sources (as reported by sequence submitters). An expressed sequence tag or E.S.T. is a short sub-sequence of a cDNA sequence that may be used to identify gene transcripts, and are instrumental in gene discovery and gene sequence determination (Expressed sequence tag, 2015). An EST results from one-shot sequencing of a cloned cDNA. The cDNAs used for EST generation are in general individual clones from a cDNA library. The resulting sequence is a relatively low quality fragment, approximately 500 to 800 nucleotides. Because these clones consist of DNA that is complementary to mRNA, the ESTs represent portions of expressed genes (Expressed sequence tag, 2015).

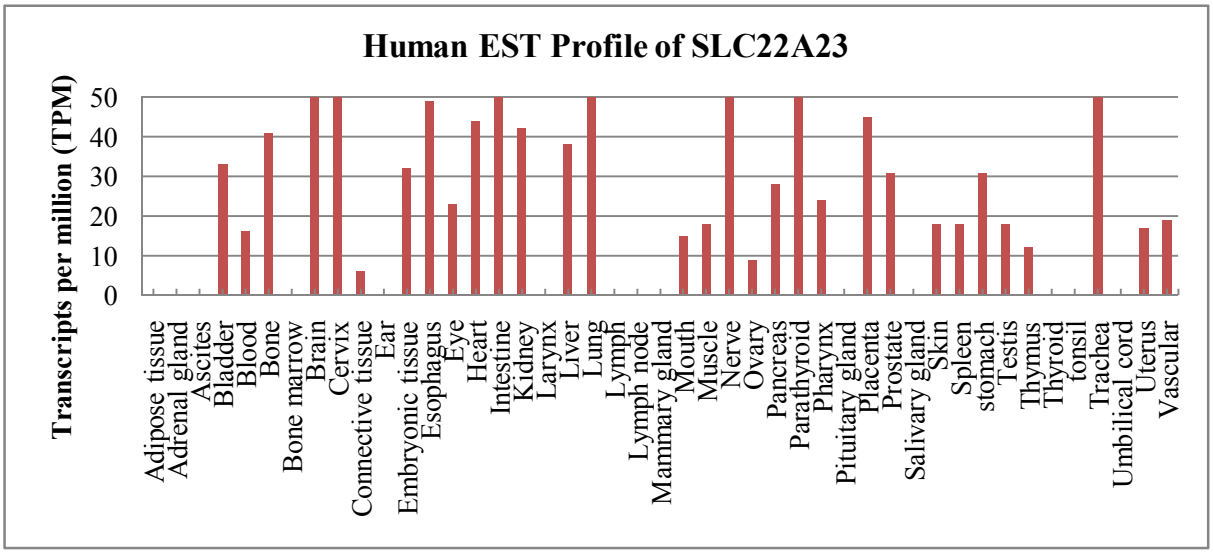


Figure 8: Human EST profile of *SLC22A23* by body site. Sourced and analyzed from UniGene (April 2015) and expressed as Transcripts per Million (TPM).

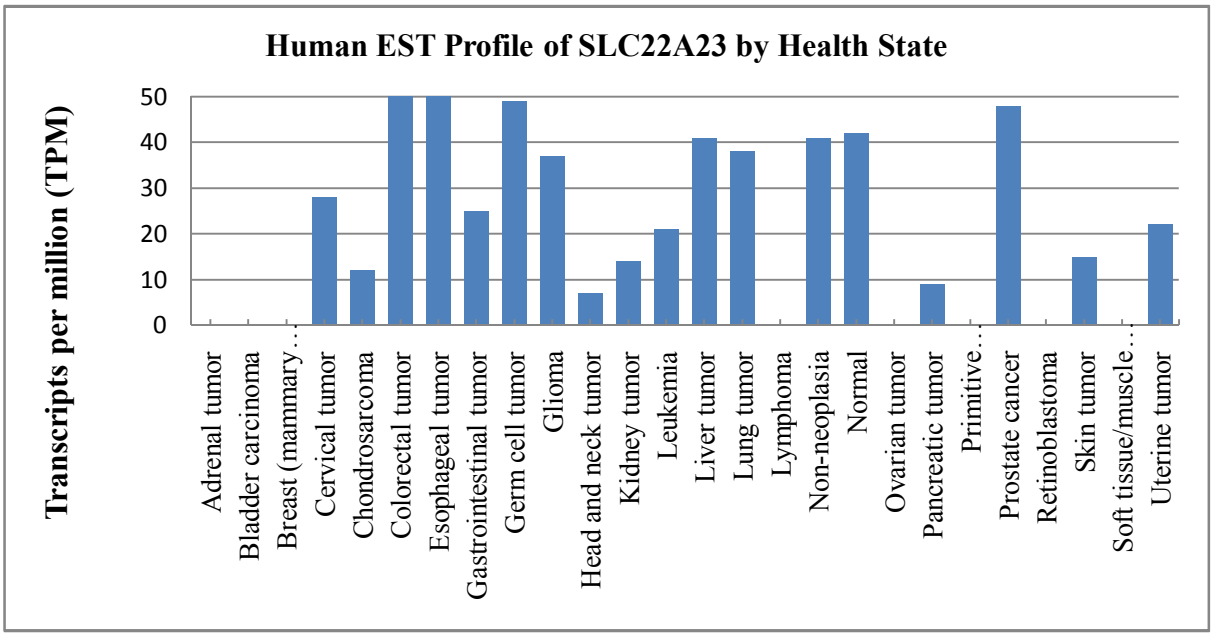


Figure 9: Human EST profile of *SLC22A23* by health state. Sourced and analyzed from UniGene (April 2015) and expressed as Transcripts per Million (TPM).

b) AceView browser:

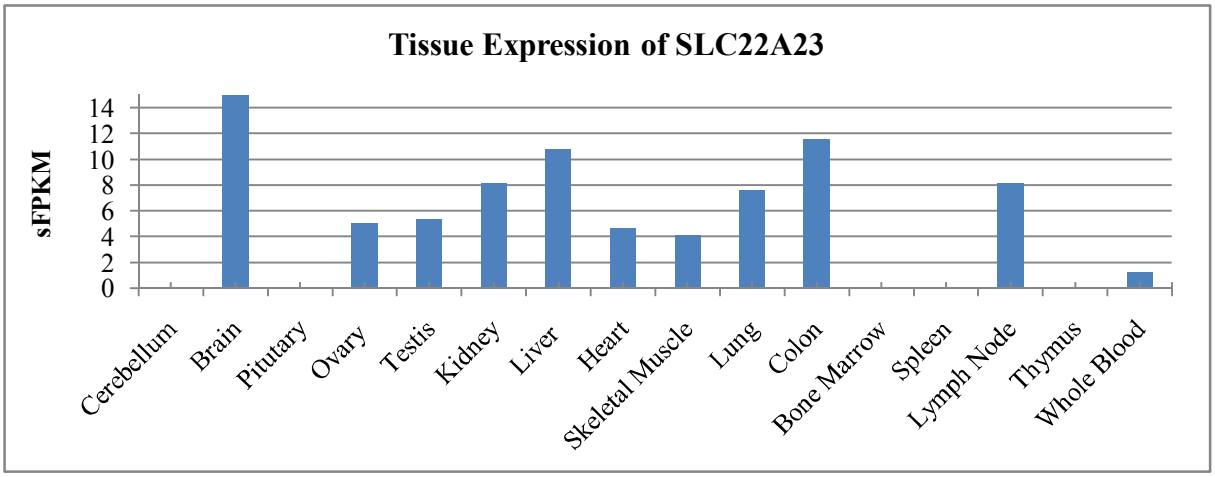


Figure 10: Tissues expression of *SLC23A23*. Tissues of significant RNA expression are determined from publicly available RNA-sequencing datasets and expressed in significant Fragments Per Kilobase Of Exon Per Million Fragments Mapped (sFPKM).

In figure 10, the abundance of the human mRNA in specific tissues were deduced from RNA sequencing data, determined by the Reference Transcriptome Resource (http://nhprtr.org/) (Pipes et al 2013). The abundance of reads was assessed by the magic index against the

intergenic background and expression values were normalized as significant fragments per kilobase of exon per million fragments mapped (sFPKM) (Pipes et al 2013).

According to UniGene expression profile and Ace View genome browser *SLC22A23* is abundantly expressed in wide variety of tissues including strong expression in brain, esophagus, intestine, lung, cervix, nerve, parathyroid, placenta, trachea, colon and liver. In disease condition, *SLC22A23* is highly expressed in colorectal tumor, esophageal tumor, germ cell tumour and prostate cancer.

7. 2. Alignment Analysis Of The Human SLC22A23 Genomic Structure

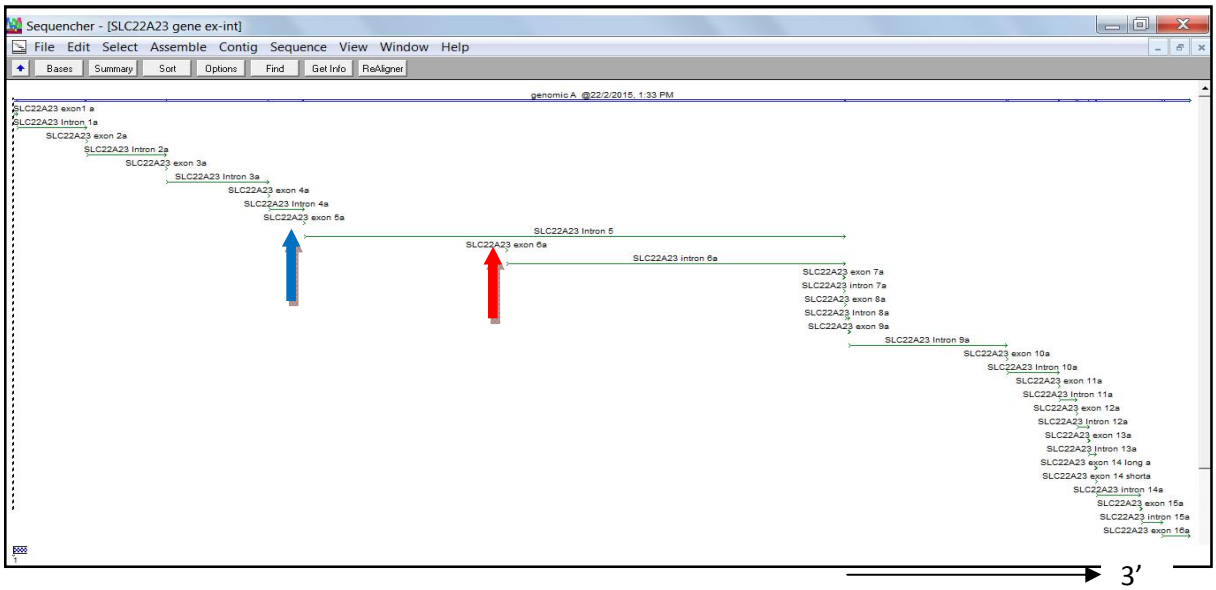


Figure 11: Alignment of the transcripts to the *SLC22A23* locus.

Figure 11 demonstrates the exon intron structure of *SLC22A23* gene after alignment of the transcripts with genomic sequence. Based on reference RNAs we found 13 exons and based on ESTs we found 3 new exons. The red arrow in the figure 12 is showing the exon 6; which has a new transcriptional start site. Blue arrow is showing exon 5 which has two alternative starting points. In total we found 16 exons and 15 corresponding introns (Figure 12).

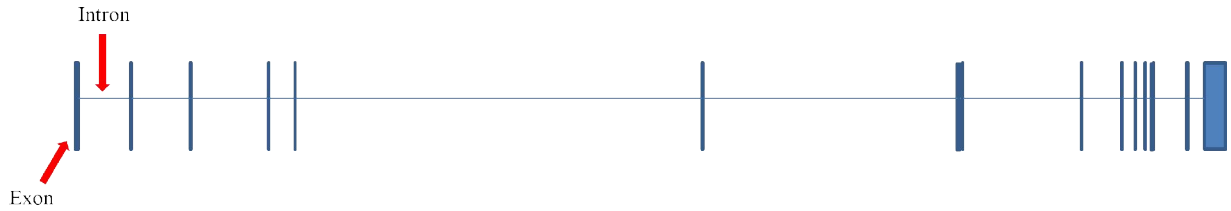


Figure 12: Structure of *SLC22A23* gene, boxes represents number of exons, line represents number of introns, and overall size represents proportion of nucleotide length.

7.3. Tissue Expression Of The Human SLC22A23

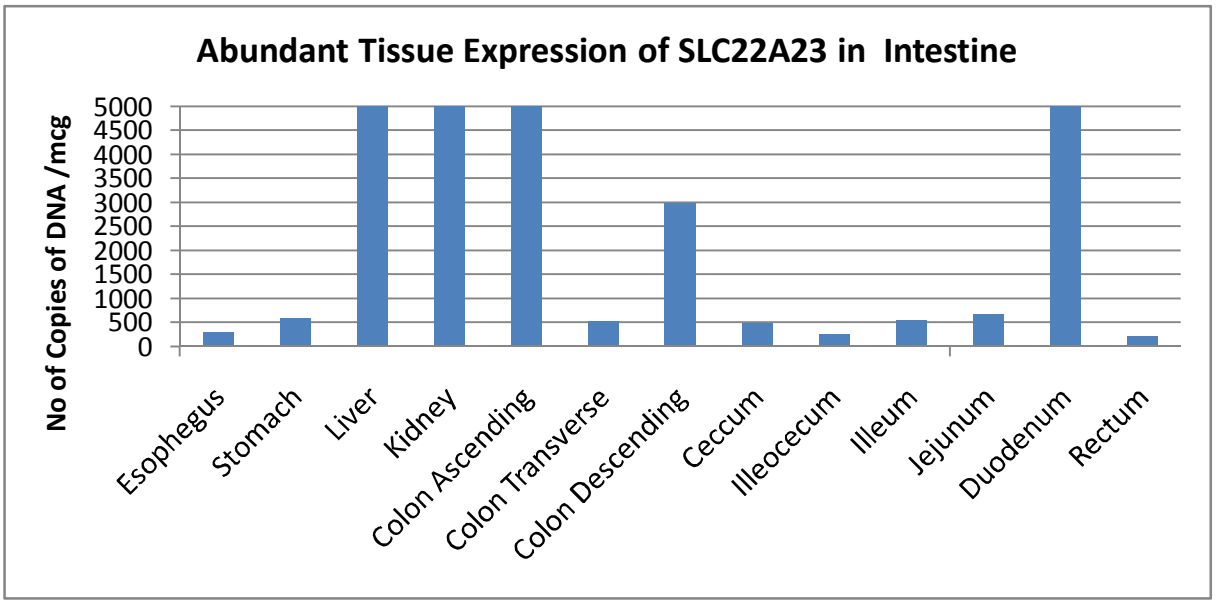


Figure 13: Tissue expression of *SLC22A23* in intestinal sections determined by Real Time PCR.

Performing tissue expression analysis, using real time PCR, we found that the human *SLC22A23* is abundantly expressed in wide range of tissues.

Figure 13 demonstrates *SLC22A23* is expressed in the intestine; with high expression in colon ascending, colon descending and colon. *SLC22A23* is also present in the brain, pancreas and lung (Figure 14).

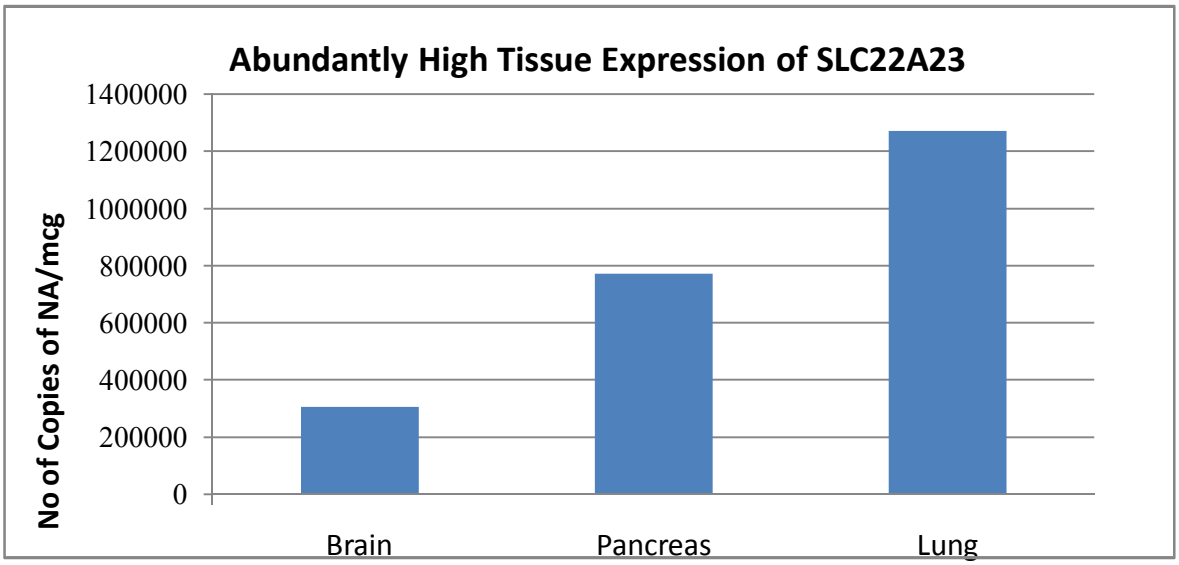


Figure 14: Tissue expression of *SLC22A23* determined by Real Time PCR.

7.4. Cellular Localization *SLC22A23*

Initially we used lipofectamine® 2000 reagent and lipofectamine® 3000 transfection reagent for all three cell lines; at concentration of 9 and 15 µL, but no fluorescent protein was observed even after increasing the reagent concentration upto 18µL. Primarily, effectene transfection reagent also did not express any fluorescent protein when used at a concentration mentioned in the protocol for 6 well plates; unless the reagent concentration was increased to 3 times. However; the transfection efficiency was very low in all three cell lines all times. After 72h of transient transfection using effectene and mirus transfection reagent tGFP fluorescent protein was observed using fluorescence microscope in CHO cell lines. It took more than 5 days to express fluorescent protein in MDCK and CaCo-2 cells. Images of live cells expressing the fluorescent protein were captured. We found that green fluorescence protein was expressed on the cell membrane of all three cell lines (Figure 15) and no green fluorescence protein was observed in the controls.

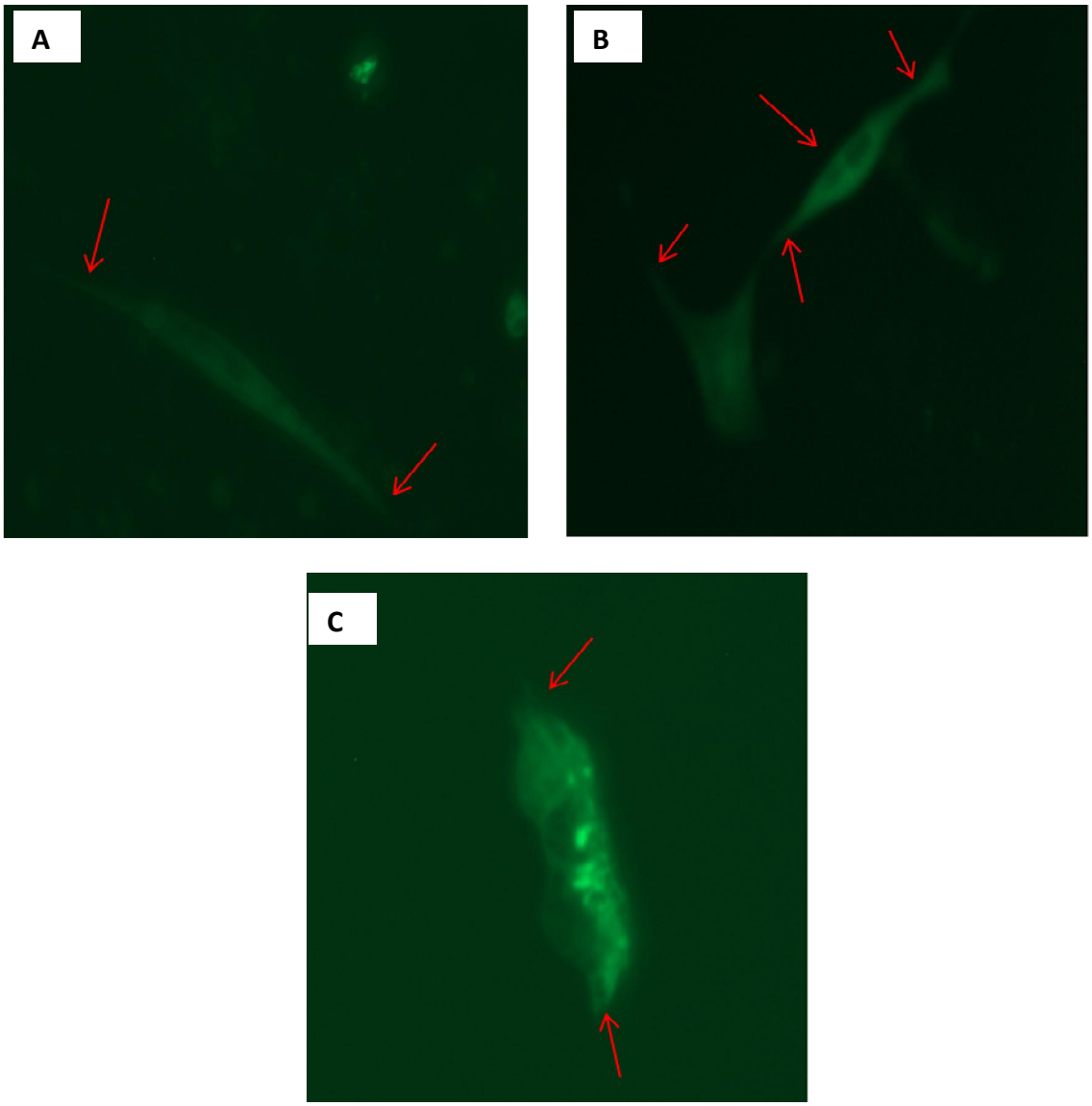


Figure 15: Expression of SLC22A23 protein in the cell membrane of CHO cell (panel A), MDCK cell (panel B) and CaCo-2 cell (panel C) observed in GFP light.

7.5. Generation Of Monoclonal Stable Cell Lines

Monoclonal stable cell lines could not be generated. To create stable cell lines, cells were transfected several times with effectene and mirus transfection reagent. The transfection efficiency was always very low.

Transient transfection showed many live fluorescent cells however, overtime number of live cells decreased to one or zero in most cases for all three cells lines. Predominantly for MDCK

and CaCo-2 cells number of live fluorescent cells became zero within 1.5-2 weeks of transfection. Live fluorescent cells did not multiply enough even after splitting and feeding the cells properly. Number of dead cells was always high in most transfections. In case of MDCK and CaCo-2 cells only 1-2 times few live fluorescent cells were observed after 2 weeks. And for CHO cells, several times some live fluorescent cells were observed after 4 weeks (stable transfection). In some cases, 2-3 CHO cell colonies were observed however; most of the time the live fluorescent cells were present in scattered way on the surface of the wells. As a result we were close enough to create some CHO stable cells that survived more than a month (Figure 16). Nevertheless, we were unable to separate these single live fluorescent cells or colonies from the non-transfected cells. In order to separate the transfected cells (either colony of single fluorescent cells) from the non-transfected ones the following methods were used for all three cell lines.

7.5.1. Dilution Method – High concentrations of geneticin, G418, were used as our medium contained serum. However; 700-1000µg/ml of G418 was unable to kill the non-transfected cells as these cells developed resistance overtime to G418. Hence the stable transfected cells could not be separated. We tried to lower the cells density and plated them in 96 well plates ensuring one cell per well. Still we were unsuccessful in creating a stable cell line.

7.5.2. Cloning Cylinder Method – In order to grow colonies we tried to separate the single fluorescent cell from the non-transfected ones and placed in a 96 well plate. However; the cells died within 24-48 hours.

7.5.3. Fluorescence Activated Cell Sorting – Overall we did monoclonal FACS three times in a 96 well plate using CHO cells and only one time with MDCK and CaCo-2 cell. In the first time, we did FACS for all three cell lines but lost them after FACS due to incubator contamination. The next two times only CHO cell was sorted where during the second time

soon after sorting we observed the cells did not multiply and died in 96 well plates. And in the third time there was no sign of cells in the 96 well plates. This might be because single sorted cells are more prone to death and do not grow very fast. Sorted cells are quite fragile as they go through pressure and electrical force and in our case the expression of the sorted cells were quite low due to very low transfection efficiency with low expression level of GFP. As a result we were unable to establish stable cell lines; expressing SLC22A23 and were unable to do the functional study.

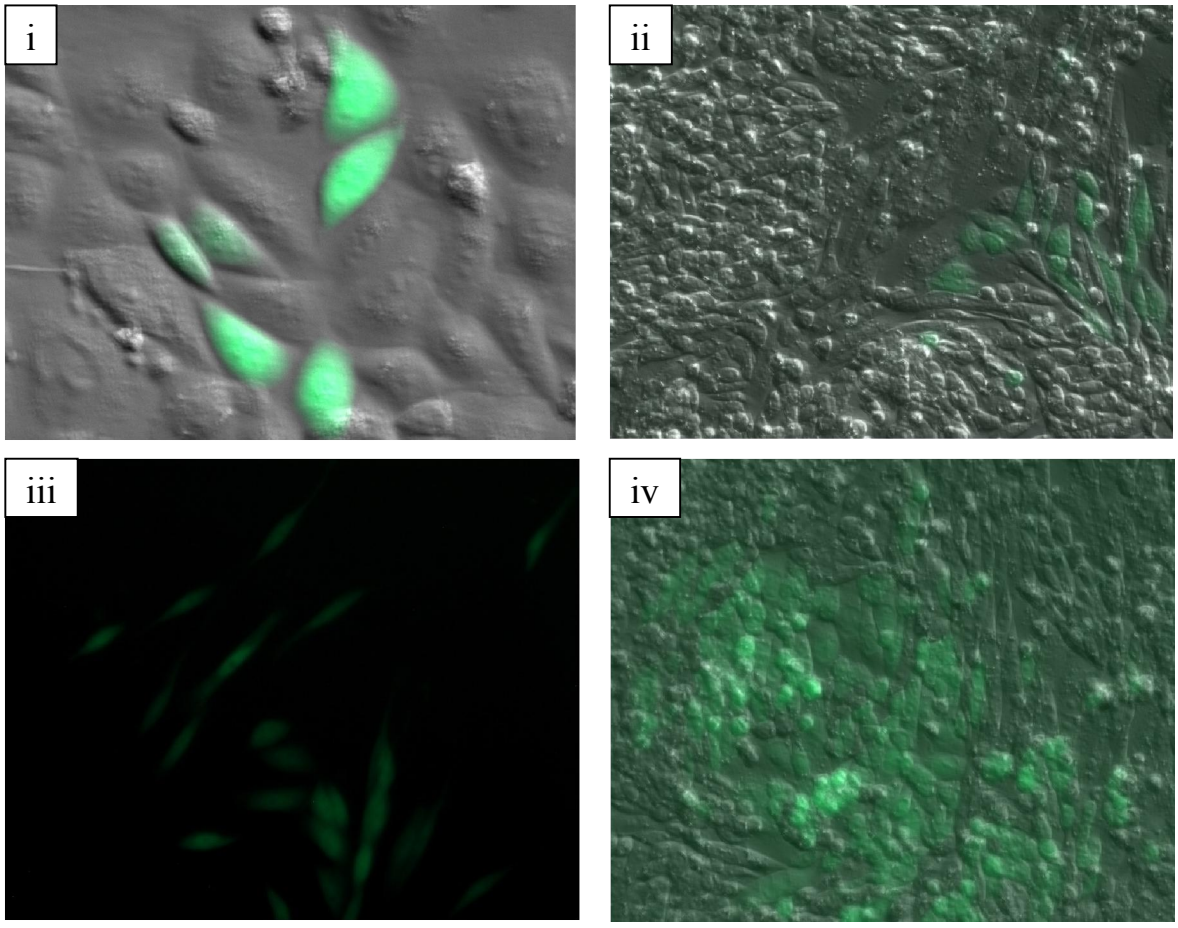


Figure 16: i) Stably transfected CHO cell colony in bright light, ii) Stably transfected CHO cell colony in bright light surrounded by non-transfected, resistant to G418 CHO cells, iii) Stably transfected CHO cell colony in GFP light & iv) High expression of stably transfected CHO cells in bright light surrounded by non-transfected, resistant to G418 CHO cells.

8. DISCUSSIONS

Being a chronic disease IBD is steadily increasing worldwide and is a common cause of gastrointestinal morbidity predominantly in the Western world (Noble et al., 2005). Significant variation in the incidence and prevalence rates of IBD between different populations indicates that genetics has a strong association in the pathogenesis of IBD (Blanchard et al., 2001).

Research in the genetic background of IBD is a rapidly growing field and in the past decade, remarkable progress has been achieved in unraveling the genetic aetiology of inflammatory bowel disease (IBD) comprising ulcerative colitis and Crohn's disease (Weersma et al., 2007).

Apart from immune genes IBD have been frequently associated with genetic mutations of organic cation transporters (OCTNs) at different levels, which include *SLC22A23*, an orphan gene encoding a putative membrane transporter protein (Jacobsson et al., 2007).

The human *SLC22A23* is a member of the solute carrier family 22 and located in chromosome 6. Very little is known about its physiological function as the genomic organization, localization and genetic variations have not been described yet. *SLC22A23* is predicted to encode for an organic cation membrane transporter protein and the human gene structure and its functions is still uncharacterized. Hence it is imperative to determine *SLC22A23's* biological functional as a means to determine which solutes are imbalanced in IBD. Therefore, the aim of the study was to characterize the gene, determine its cellular localization to deduce if SLC22A23 indeed could function as a solute carrier.

Here, for the first time we have comprehensively described the chromosomal location, the gene neighborhood and the genomic structure of human *SLC22A23*. This will aid the prioritization of genetic variations in the locus in future studies.

Importantly, in contrast to the *SLC22A4* and *SLC22A5* genes in the IBD5 locus, *SLC22A23* is not located near any inflammatory genes, strengthening our hypothesis that an imbalance in organic cations could contribute to the development of IBD.

Our analysis confirmed that *SLC22A23* has 16 exons, whereas according to the databases it has 11 to 15 exons. The human EST profile showed that the gene is expressed in wide variety of tissues. *SLC22A23* is abundantly expressed in brain, esophagus, intestine and lung. Similar tissue expression was reported for the rat homolog BOCT2 (*slc22a23*) by Bennett et al. in 2011 (Bennett et al., 2011). Based on health state, the gene is highly expressed in colorectal tumor, esophageal tumor and gastrointestinal tumor; which supports the fact *SLC22A23* is involved in the pathogenesis of IBD, since the inflammation is precursory to cancer development (Mattar et al., 2011).

To understand the biology of the *SLC22A23*, we subcloned and expressed the human *SLC22A23* gene that showed plasma membrane localization (Figure 15), confirming that the human *SLC22A23* gene encodes a membrane protein which indicated that it is involved in solutes transport. Similar membrane localization was reported for the rat homolog BOCT2, but transport studies on the rat homolog failed to identify the substrate.

For the first time we performed tissue expression of human *SLC22A23* and we found the gene expressed in wide variety of tissues and abundantly expressed in brain, lung, pancreas and in the intestine. Expression of the *SLC22A23* in the intestinal tissues strengthens our hypothesis that *SLC22A23* is significant to IBD.

In order to perform functional studies we tried to create a monoclonal cell line using different methods. However, due to contamination of cells, antibiotic resistance of non transfected cells and low transfection efficiency we were unsuccessful to establish a stable cell lines expressing *SLC22A23*.

9. CONCLUSIONS

Our study reports the genomic structure, cell localization and tissue expression of the human *SLC22A23* gene. The study showed the gene is not closely located to immune genes. Correct sub-cellular location confirms that the human *SLC22A23* is a membrane transporter protein (as known from the rat orthologous). And the human *SLC22A23* is expressed in the intestine which is relevant for IBD.

This research will serve as the basis for doing the functional study of *SLC22A23*, which would be a crucial step as the substrate(s) will lead to the mechanism in the development of IBD and will contribute to the understanding whether metabolite or nutrient imbalance plays a role in the development of IBD.

More in depth characterization of the *SLC22A23* gene will provide the basis to determine which genetic variation is causative in the development of IBD in future studies. Functional variation contributing to IBD can later be verified in biological assays. And eventually, functional foods or nutritional supplements can be developed to prevent intestinal inflammation in individuals carrying a higher risk for intestinal inflammation.

10. APPENDIX

10.1. Genetic variant found in the *SLC22A23* coding region

a) Splice Donar Variants (7)

ID	Chr: bp	Alle-les	Glo-bal MAF	Class	Sou rce	Type	AA	AA coord	SIFT	Poly-Phen	Tran script
rs151086698	6:3286858	C/G	-	SNP	dbSNP	Splice donor variant NMD transcript variant	-	-	-	-	ENST00000497691
rs369575802	6:3289763	C/T	-	SNP	dbSNP	Splice donor variant	-	-	-	-	ENST00000380302
rs73722419	6:3323264	A/T	0.056 (T)	SNP	dbSNP	Splice donor variant NMD transcript variant	-	-	-	-	ENST00000497691
COSM405214	6:34101157-34101190	34/-	-	somatic_sequence_alteration	COSMIC	Splice donor variant Coding sequence variant Intron variant Feature truncati		304	-	-	ENST00000380298

						on					
COSM405215	6:34100157-34101190	34/-	-	somatic_sequence_alteration	COSMIC	Splice donor variant Coding sequence variant Intron variant Feature truncation		304	-	-	ENST00000380298
rs188995683	6:343 8483	C/T	0.000 (T)	SNP	dbSNP	Splice donor variant	-	-	-	-	ENST00000467177
rs148146560	6:344 4818	C/T	0.000 (T)	SNP	dbSNP	Splice donor variant	-	-	-	-	ENST00000380302

b) Stop Gained (6)

ID	Chr: bp	Alle-les	Glo-bal MAF	Class	Sou rce	Type	AA	AA coord	SIFT	Poly-Phen	Tran script
COSM3777545	6:328 5081	G/C	-	somatic_SNV	COSMIC	Stop gained Splice region variant	S/*	245	-	-	ENST00000380302
COSM3777546	6:328 5081	G/C	-	somatic_SNV	COSMIC	Stop gained Splice region variant	S/*	245	-	-	ENST00000380302
COSM3625541	6:328 9764	G/T	-	somatic_SNV	COSMIC	Stop gained Splice	S/*	157	-	-	ENST00000380302

						region variant					2
COSM3625542	6:3289764	G/T	-	somatic_SNV	COSMIC	Stop gained Splice region variant	S/*	157	-	-	ENST00000380302
COSM3873121	6:3323890	C/T	-	somatic_SNV	COSMIC	Stop gained	W/*	342	-	-	ENST00000380298
COSM3873122	6:3323890	C/T	-	somatic_SNV	COSMIC	Stop gained	W/*	342	-	-	ENST00000380298

c) Splice Acceptor Variant (1)

ID	Chr: bp	Alle-les	Glo-bal MAF	Class	Sou rce	Type	AA	AA coord	SIFT	Poly-Phen	Tran script
COSM315259	6:3324004	T/G	-	somatic_SNV	COSMIC	Splice acceptor variant	-	-	-	-	ENST00000380298

d) Frameshift Variant (4)

ID	Chr: bp	Alle-les	Glo-bal MAF	Class	Sou rce	Type	AA	AA coord	SIFT	Poly-Phen	Tran script
TMP_ESP_6_3273330_3273331	6:3273096-3273097	GT/-	-	deletion	ESP	Frameshift variant Missense variant Feature truncation	TL/TX	392	-	-	ENST00000380302
rs34106449	6:between 3273215&	-/C	-	insertion	dbSNP	Frameshift variant Feature	E/GX	353	-	-	ENST00000380302

	3273216					elongation					
rs35544863	6: between 3455920 & 3455921	-/A	-	insertion	dbSNP	Frameshift variant Feature elongation	-/X	214	-	-	ENST00000380298
COSM1621551	6:3456452	C/-	-	somatic_deletion	COSMIC	Frameshift variant Missense variant Feature truncation	A/X	36	-	-	ENST00000380298

e) Missense Variant (128)

ID	Chr: bp	Alleles	Global MAF	Class	Source	Type	AA	AA coord	SIFT	PolyPhen	Transcript
COSM1077879	6:3273072	C/T	-	somatic_SNV	COSMIC	Missense variant	G/S	401	0.05	0.067	ENST00000380302
COSM1077880	6:3273072	C/T	-	somatic_SNV	COSMIC	Missense variant	G/S	401	0.05	0.067	ENST00000380302
rs37296311 2	6:3273075	T/C	-	SNP	dbSNP	Missense variant	N/D	400	0.02	0.785	ENST00000380302
TMP_ESP_6_3273330_32733331	6:3273096-3273097	GT/-	-	deletion	ESP	Frameshift variant Missense variant Feature	TL/TX	392	-	-	ENST00000380302

						truncation					
rs11206542 9	6:32731 05	C/T	-	SNP	dbSNP	Missense variant	G/S	390	0.56	0.007	ENST00000380302
COSM3873000	6:32731 07	G/A	-	somatic_SNV	COSMIC	Missense variant	A/V	389	0.55	0	ENST00000380302
COSM3873001	6:32731 07	G/A	-	somatic_SNV	COSMIC	Missense variant	A/V	389	0.55	0	ENST00000380302
rs15044472 7	6:32731 08	C/T	-	SNP	dbSNP	Missense variant	A/T	389	0.13	0.002	ENST00000380302
rs2013965 52	6:32731 20	C/T	0.000 (T)	SNP	dbSNP	Missense variant	D/N	385	0.04	0.595	ENST00000490273
rs3692578 93	6:32731 28	C/G	-	SNP	dbSNP	Missense variant	G/A	382	0.62	0.913	ENST00000380302
rs3728070 40	6:32731 31	G/A	-	SNP	dbSNP	Missense variant	S/L	381	0	0.876	ENST00000490273
COSM3873002	6:32731 47	C/T	-	somatic_SNV	COSMIC	Missense variant	E/K	376	0.25	0.362	ENST00000380302
COSM3873003	6:32731 47	C/T	-	somatic_SNV	COSMIC	Missense variant	E/K	376	0.25	0.362	ENST00000380302
rs3711289 49	6:32731 77	C/T	-	SNP	dbSNP	Missense variant	G/R	366	0.07	0.609	ENST00000490273

rs141223516	6:32731185	T/G	-	SNP	dbSNP	Missense variant	H/P	363	0.29	0.001	ENST00000380302
COSM4161038	6:32732203	C/G	-	somatic_SNV	COSMIC	Missense variant	R/P	357	0	0.988	ENST00000380302
COSM4161039	6:32732203	C/G	-	somatic_SNV	COSMIC	Missense variant	R/P	357	0	0.988	ENST00000380302
rs199592727	6:32732204	G/A	-	SNP	dbSNP	Missense variant	R/C	357	0.002	0.429	ENST00000380302
COSM3873004	6:32732204	G/A	-	somatic_SNV	COSMIC	Missense variant	R/C	357	0.002	0.429	ENST00000380302
COSM3873005	6:32732204	G/A	-	somatic_SNV	COSMIC	Missense variant	R/C	357	0.002	0.429	ENST00000380302
COSM1186937	6:32732207	T/C	-	somatic_SNV	COSMIC	Missense variant	T/A	356	0.24	0.41	ENST00000380302
COSM1186938	6:32732207	T/C	-	somatic_SNV	COSMIC	Missense variant	T/A	356	0.24	0.41	ENST00000380302
rs201540017	6:32732222	T/C	-	SNP	dbSNP	Missense variant	N/D	351	1	0.002	ENST00000380302
COSM3625491	6:32732236	G/A	-	somatic_SNV	COSMIC	Missense variant	P/L	346	0	0.999	ENST00000380302
COSM3625492	6:32732236	G/A	-	somatic_SNV	COSMIC	Missense variant	P/L	346	0	0.999	ENST00000380302

COSM376408	6:32732745	T/C	-	somatic_SNV	COSMIC	Missense variant	Q/R	343	0.12	0.899	ENST00000380302
COSM376409	6:32732745	T/C	-	somatic_SNV	COSMIC	Missense variant	Q/R	343	0.12	0.899	ENST00000380302
COSM1696867	6:32732758	C/T	-	somatic_SNV	COSMIC	Missense variant	E/K	339	0	1	ENST00000380302
COSM1696868	6:32732758	C/T	-	somatic_SNV	COSMIC	Missense variant	E/K	339	0	1	ENST00000380302
COSM1443562	6:32733327	A/G	-	somatic_SNV	COSMIC	Missense variant	Y/H	316	0	0.999	ENST00000380302
COSM1443563	6:32733327	A/G	-	somatic_SNV	COSMIC	Missense variant	Y/H	316	0	0.999	ENST00000380302
COSM1443561	6:32733327	A/T	-	somatic_SNV	COSMIC	Missense variant	Y/N	316	0	0.999	ENST00000380302
COSM1443560	6:32733327	A/T	-	somatic_SNV	COSMIC	Missense variant	Y/N	597	0	0.999	ENST00000406686
COSM3777538	6:32733348	C/T	-	somatic_SNV	COSMIC	Missense variant	E/K	309	0.001	0.801	ENST00000380302
COSM3777539	6:32733348	C/T	-	somatic_SNV	COSMIC	Missense variant	E/K	309	0.001	0.801	ENST00000380302
rs148262614	6:32733357	G/A	-	SNP	dbSNP	Missense variant	P/S	306	0.18	1	ENST00000380302

rs200011775	6:32733381	C/T	-	SNP	dbSNP	Missense variant	A/T	298	0.01	1	ENST00000380302
COSM1226041	6:32734405	C/T	-	somatic_SNV	COSMIC	Missense variant	G/R	290	0	1	ENST00000380302
COSM1226042	6:32734405	C/T	-	somatic_SNV	COSMIC	Missense variant	G/R	290	0	1	ENST00000380302
rs37702816 3	6:32837326	C/T	-	SNP	dbSNP	Missense variant	R/H	438	-	0	ENST00000485307
rs11724447 47	6:32837350	A/G	0.006 (G)	SNP	dbSNP	Missense variant	L/P	430	-	0	ENST00000485307
rs200940909	6:32838320	C/T	0.001 (T)	SNP	dbSNP	Missense variant	G/S	407	0	1	ENST00000485307
rs182163436	6:32838368	T/C	0.001 (C)	SNP	dbSNP	Missense variant	T/A	282	0.31	0.679	ENST00000380302
rs37702816 3	6:32837326	C/T	-	SNP	dbSNP	Missense variant	R/H	438	-	0	ENST00000485307
rs199958017	6:32838369	G/C	-	SNP	dbSNP	Missense variant	I/M	281	0	0.992	ENST00000490273
rs202186678	6:32839321	A/G	-	SNP	dbSNP	Missense variant	M/T	264	0.07	0.876	ENST00000380302
rs371116667	6:32839339	G/A	-	SNP	dbSNP	Missense variant	A/V	258	0.57	0.21	ENST00000380302

rs149762554	6:32839340	C/T	-	SNP	dbSNP	Missense variant	A/T	258	1	0.02	ENST00000490273
COSM1634745	6:32839346	A/T	-	somatic_SNV	COSMIC	Missense variant	S/T	256	0.81	0.557	ENST00000380302
COSM1634746	6:32839346	A/T	-	somatic_SNV	COSMIC	Missense variant	S/T	256	0.81	0.557	ENST00000380302
rs145712961	6:32839361	C/G	-	SNP	dbSNP	Missense variant	V/L	251	1	0.017	ENST00000380302
COSM1077907	6:32839371	C/T	-	somatic_SNV	COSMIC	Missense variant	M/I	247	0.33	0.03	ENST00000380302
COSM1077908	6:32839371	C/T	-	somatic_SNV	COSMIC	Missense variant	M/I	247	0.33	0.03	ENST00000380302
rs148573369	6:32850083	G/T	-	SNP	dbSNP	Missense variant	D/E	244	0.14	0.05	ENST00000490273
rs201461663	6:32850088	G/A	-	SNP	dbSNP	Missense variant	P/S	243	0.29	0.926	ENST00000380302
COSM1077909	6:32869310	A/C	-	somatic_SNV	COSMIC	Missense variant	F/V	218	0.52	0.773	ENST00000380302
COSM1077910	6:32869310	A/C	-	somatic_SNV	COSMIC	Missense variant	F/V	218	0.52	0.773	ENST00000380302
COSM3873039	6:32869352	C/T	-	somatic_SNV	COSMIC	Missense variant	V/M	204	0.12	0.202	ENST00000380302

COSM3873040	6:32869052	C/T	-	somatic_SNV	COSMIC	Missense variant	V/M	204	0.12	0.202	ENST00000380302
rs37035129 2	6:32869058	T/C	-	SNP	dbSNP	Missense variant	M/V	202	0.28	0.004	ENST00000380302
rs200186759	6:32869066	C/T	0.000 (T)	SNP	dbSNP	Missense variant	C/Y	199	0	1	ENST00000380302
COSM1226045	6:32869069	G/A	-	somatic_SNV	COSMIC	Missense variant	S/F	198	0	0.992	ENST00000380302
COSM1226046	6:32869069	G/A	-	somatic_SNV	COSMIC	Missense variant	S/F	198	0	0.992	ENST00000380302
rs150075690	6:32869090	G/A	-	SNP	dbSNP	Missense variant	T/M	191	0.68	0.003	ENST00000490273
COSM741721	6:32870003	C/T	-	somatic_SNV	COSMIC	Missense variant	D/N	187	0.31	0.082	ENST00000380302
COSM741722	6:32870003	C/T	-	somatic_SNV	COSMIC	Missense variant	D/N	187	0.31	0.082	ENST00000380302
rs138414914	6:32870017	T/A	-	SNP	dbSNP	Missense variant	E/V	182	0.21	0.173	ENST00000490273
rs140418737	6:32870026	G/A	0.001 (A)	SNP	dbSNP	Missense variant	P/L	179	0.01	0.322	ENST00000380302
COSM1443611	6:32870078	C/T	-	somatic_SNV	COSMIC	Missense variant	G/R	162	0	0.999	ENST00000380302

COSM1443612	6:3287078	C/T	-	somatic_SNV	COSMIC	Missense variant	G/R	162	0	0.999	ENST00000380302
COSM315260	6:3289786	T/C	-	somatic_SNV	COSMIC	Missense variant	I/V	431	0.1	0.708	ENST00000436008
COSM315261	6:3289786	T/C	-	somatic_SNV	COSMIC	Missense variant	I/V	257	0.3	0.872	ENST00000467177
rs20010445 3	6:3289804	G/A	0.0000 (A)	SNP	dbSNP	Missense variant	R/W	144	0	0.997	ENST00000490273
rs14005489 9	6:3289833	T/A	-	SNP	dbSNP	Missense variant	K/M	134	0.1	0.984	ENST00000380302
rs14146896 7	6:3289834	T/C	-	SNP	dbSNP	Missense variant	K/E	134	0.23	0.867	ENST00000490273
rs37641850 4	6:3289848	G/A	-	SNP	dbSNP	Missense variant	S/F	129	0.76	0.492	ENST00000380302
COSM373690	6:3298107	G/C	-	somatic_SNV	COSMIC	Missense variant	I/M	117	0.32	0.365	ENST00000380302
COSM373691	6:3298107	G/C	-	somatic_SNV	COSMIC	Missense variant	I/M	117	0.32	0.365	ENST00000380302
rs371110850	6:3298121	G/C	-	SNP	dbSNP	Missense variant	P/A	113	0.52	0.012	ENST00000490273
rs18406434 8	6:3298136	T/G	0.0000 (G)	SNP	dbSNP	Missense variant	K/Q	108	0.02	0.889	ENST00000380302

COSM1077938	6:32981184	T/C	-	somatic_SNV	COSMIC	Missense variant	T/A	92	0	0.953	ENST00000380302
COSM1077939	6:32981184	T/C	-	somatic_SNV	COSMIC	Missense variant	T/A	92	0	0.953	ENST00000380302
rs374369262	6:32981188	C/T	-	SNP	dbSNP	Missense variant	M/I	90	0.29	0.055	ENST00000490273
COSM451286	6:32982210	G/A	-	somatic_SNV	COSMIC	Missense variant	P/L	83	0.34	0.988	ENST00000380302
COSM451287	6:32982210	G/A	-	somatic_SNV	COSMIC	Missense variant	P/L	83	0.34	0.988	ENST00000380302
COSM292459	6:33238234	G/A	-	somatic_SNV	COSMIC	Missense variant	S/L	361	0	0.022	ENST00000380298
COSM292458	6:33238234	G/A	-	somatic_SNV	COSMIC	Missense variant	S/L	361	0	0.022	ENST00000380298
COSM1226047	6:33238240	T/C	-	somatic_SNV	COSMIC	Missense variant	Y/C	359	0.08	0.971	ENST00000380298
COSM1226048	6:33238240	T/C	-	somatic_SNV	COSMIC	Missense variant	Y/C	359	0.08	0.971	ENST00000380298
rs201802673	6:33238249	A/T	0.0000 (T)	SNP	dbSNP	Missense variant	M/K	356	0	0.628	ENST00000380298
COSM285042	6:33238252	A/G	-	somatic_SNV	COSMIC	Missense variant	L/P	355	0	0.098	ENST00000380298

COSM285041	6:33238352	A/G	-	somatic_SNV	COSMIC	Missense variant	L/P	355	0	0.098	ENST00000380298
COSM1546584	6:33238379	T/A	-	somatic_SNV	COSMIC	Missense variant	Q/L	346	0	0.988	ENST00000380298
COSM1546585	6:33238379	T/A	-	somatic_SNV	COSMIC	Missense variant	Q/L	346	0	0.988	ENST00000380298
rs369931367	6:33239306	G/A	-	SNP	dbSNP	Missense variant	A/V	337	0.5	0.049	ENST00000380298
rs2010502070	6:33239307	C/T	0.001 (T)	SNP	dbSNP	Missense variant	A/T	337	0.18	0.509	ENST00000380298
COSM742358	6:33239307	C/T	-	somatic_SNV	COSMIC	Missense variant	A/T	337	0.18	0.509	ENST00000380298
COSM742359	6:33239307	C/T	-	somatic_SNV	COSMIC	Missense variant	A/T	337	0.18	0.509	ENST00000380298
rs372180989	6:33239325	G/A	-	SNP	dbSNP	Missense variant	L/F	331	0.01	0.182	ENST00000380298
rs144784887	6:33239330	T/C	-	SNP	dbSNP	Missense variant	Q/R	329	0.008	0.225	ENST00000380298
rs200029441	6:33239336	G/A/C	-	SNP	dbSNP	Missense variant	A/G	327	0.35	0.037	ENST00000380298
COSM1226043	6:33239336	G/A	-	somatic_SNV	COSMIC	Missense variant	A/V	327	0.27	0.021	ENST00000380298

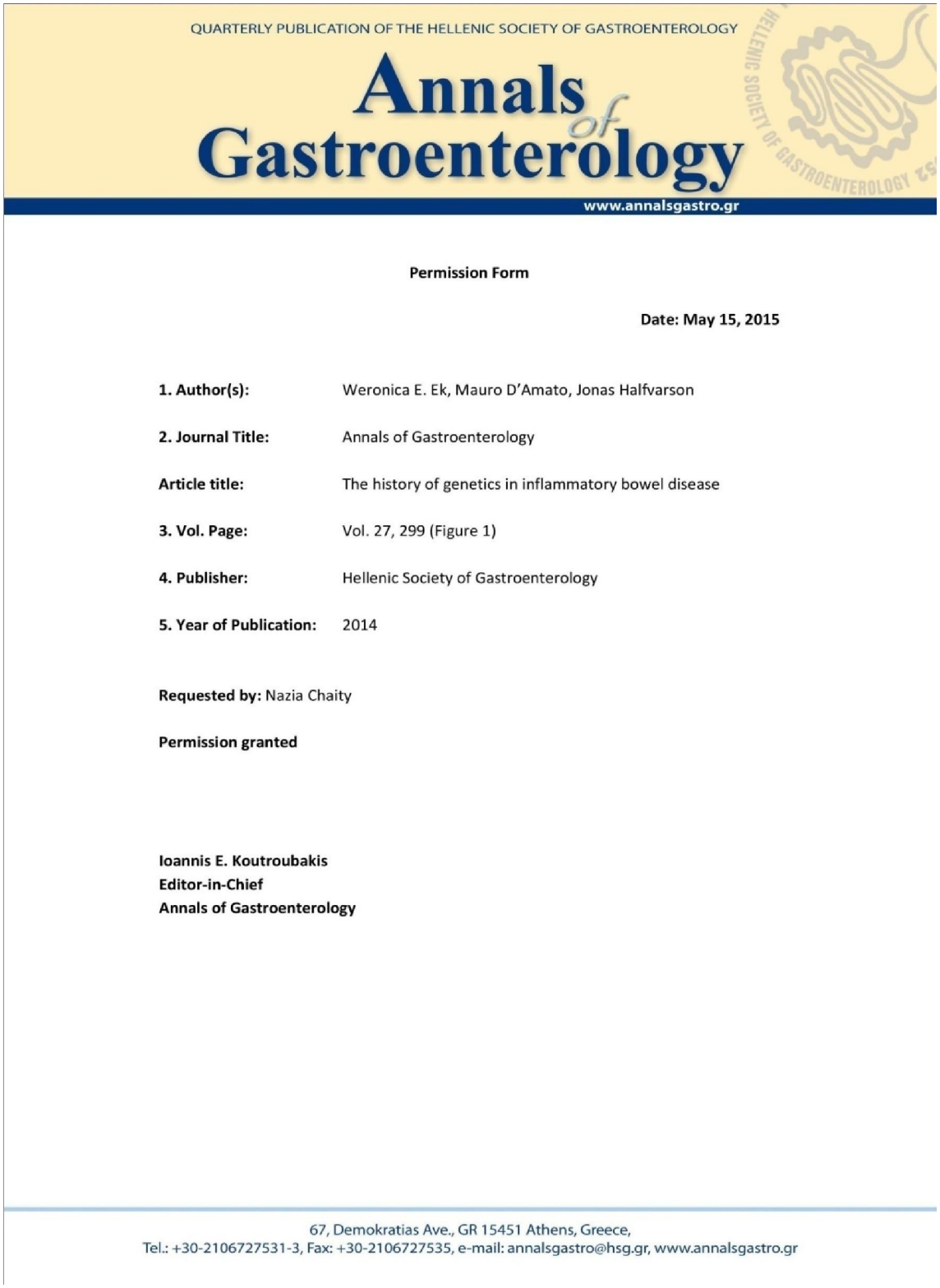
COSM1226044	6:3323936	G/A	-	somatic_SNV	COSMIC	Missense variant	A/V	327	0.27	0.021	ENST00000380298
rs36892836 2	6:3323940	T/C	-	SNP	dbSNP	Missense variant	M/V	326	0.65	0.004	ENST00000380298
COSM3873123	6:3323946	C/T	-	somatic_SNV	COSMIC	Missense variant	V/M	324	0.11	0.322	ENST00000380298
COSM3873124	6:3323946	C/T	-	somatic_SNV	COSMIC	Missense variant	V/M	324	0.11	0.322	ENST00000380298
rs373444198	6:3323961	T/C	-	SNP	dbSNP	Missense variant	M/V	319	0.01	0.349	ENST00000380298
rs143100651	6:3323963	G/A	-	SNP	dbSNP	Missense variant	T/M	318	0.1	0.321	ENST00000380298
rs376011377	6:3323976	G/A	-	SNP	dbSNP	Missense variant	R/W	314	0	0.999	ENST00000380298
rs370166164	6:3410206	G/C	-	SNP	dbSNP	Missense variant	L/V	299	0.19	0.968	ENST00000380298
COSM3994892	6:3410238	A/T	-	somatic_SNV	COSMIC	Missense variant	F/Y	288	0	0.999	ENST00000380298
COSM3994893	6:3410238	A/T	-	somatic_SNV	COSMIC	Missense variant	F/Y	288	0	0.999	ENST00000380298
rs374183830	6:3410239	A/C	-	SNP	dbSNP	Missense variant	F/V	288	0.01	0.998	ENST00000380298

COSM1443830	6:34102281	C/T	-	somatic_SNV	COSMIC	Missense variant	V/M	274	0	0.999	ENST00000380298
COSM219224	6:34103335	G/A	-	somatic_SNV	COSMIC	Missense variant	R/W	256	0	1	ENST00000380298
rs18286483 6	6:34157586	A/G	0.000 (G)	SNP	dbSNP	Missense variant	F/L	242	0.24	0.892	ENST00000380298
rs7933689 8	6:34385403	C/T	0.001 (T)	SNP	dbSNP	Missense variant	R/H	38	0.16	0.001	ENST00000467177
rs1219768 9	6:34385404	G/A	0.398 (G)	SNP	dbSNP	Missense variant	R/C	38	0.05	0.003	ENST00000467177
rs1428364 38	6:34385405	C/A	0.002 (A)	SNP	dbSNP	Missense variant	K/N	37	0.14	0	ENST00000467177
COSM1078237	6:34560404	G/T	-	somatic_SNV	COSMIC	Missense variant	P/T	186	0.007	0.001	ENST00000380298
COSM1186939	6:34560434	C/T	-	somatic_SNV	COSMIC	Missense variant	G/S	176	0.76	0.003	ENST00000380298
rs375868402	6:34560438	G/T	-	SNP	dbSNP	Missense variant	S/R	2	0.001	0.023	ENST00000485307
rs1484902 36	6:34561400	T/C	0.004 (C)	SNP	dbSNP	Missense variant	S/G	154	0.4	0.001	ENST00000380298
COSM3782057	6:34561433	T/A	-	somatic_SNV	COSMIC	Missense variant	T/S	143	0.67	0.002	ENST00000380298

COSM3697803	6:34561772	C/T	-	somatic_SNV	COSMIC	Missense variant	G/R	130	0.52	0.016	ENST00000436008
rs199852691	6:34561996	C/T	-	SNP	dbSNP	Missense variant	D/N	122	0.5	0.532	ENST00000380298
COSM3782058	6:34562992	C/T	-	somatic_SNV	COSMIC	Missense variant	G/S	90	0.01	1	ENST00000436008
COSM1621551	6:34564452	C/-	-	somatic_deletion	COSMIC	Frameshift variant Missense variant Feature truncation	A/X	36	-	-	ENST00000380298

10.2 Permission for the Copyrighted Material

a) For Figure 1



QUARTERLY PUBLICATION OF THE HELLENIC SOCIETY OF GASTROENTEROLOGY

Annals of Gastroenterology

www.annalsgastro.gr

Permission Form

Date: May 15, 2015

1. Author(s):	Weronica E. Ek, Mauro D'Amato, Jonas Halfvarson
2. Journal Title:	Annals of Gastroenterology
Article title:	The history of genetics in inflammatory bowel disease
3. Vol. Page:	Vol. 27, 299 (Figure 1)
4. Publisher:	Hellenic Society of Gastroenterology
5. Year of Publication:	2014

Requested by: Nazia Chaity

Permission granted

Ioannis E. Koutroubakis
Editor-in-Chief
Annals of Gastroenterology

67, Demokratias Ave., GR 15451 Athens, Greece,
Tel.: +30-2106727531-3, Fax: +30-2106727535, e-mail: annalsgastro@hsg.gr, www.annalsgastro.gr

b) For Figure 2

NATURE PUBLISHING GROUP LICENSE
TERMS AND CONDITIONS

Jul 07, 2015

This is a License Agreement between Nazia A Chaity ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

License Number	3663750610130
License date	Jul 07, 2015
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Genetics
Licensed content title	High-resolution haplotype structure in the human genome
Licensed content author	Mark J. Daly, John D. Rioux, Stephen F. Schaffner, Thomas J. Hudson, Eric S. Lander
Licensed content date	Oct 1, 2001
Volume number	29
Issue number	2
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Figures	Figure 2. Block-like haplotype diversity at 5q31.
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Genetics and disease associations of organic cation transporters with IBD – special emphasis on genetic and functional studies of SLC22A23
Expected completion date	Jul 2015
Estimated size (number of pages)	100
Total	0.00 CAD

REFERENCES

Aberg, K., Adkins, D. E., Liu, Y., McClay, J. L., Bukszar, J., Jia, P., . . . van den Oord, E. J. (2012). Genome-wide association study of antipsychotic-induced QTc interval prolongation.*Pharmacogenomics J, 12*(2), 165-172. doi: 10.1038/tpj.2010.76

Allchin, W. H. (1909). A Discussion on "Ulcerative Colitis": Introductory Address. Proceedings of the Royal Society of Medicine, 2(Med Sect), 59–75.

Anderson, C. A., Massey, D. C., Barrett, J. C., Prescott, N. J., Tremelling, M., Fisher, S. A., . . . Parkes, M. (2009). Investigation of Crohn's disease risk loci in ulcerative colitis further defines their molecular relationship. *Gastroenterology, 136*(2), 523-529 e523.doi: 10.1053/j.gastro.2008.10.032

Andersson, R. E., Olaison, G., Tysk, C., & Ekbom, A. (2001).Appendectomy and protection against ulcerative colitis.*N Engl J Med, 344*(11), 808-814. doi: 10.1056/nejm200103153441104

Armuzzi, A., Ahmad, T., Ling, K.-L., de Silva, A., Cullen, S., van Heel, D., … Jewell, D. P. (2003). Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31.Gut, 52(8), 1133–1139.

Babusukumar, U., Wang, T., McGuire, E., Broeckel, U., & Kugathasan, S. (2006). Contribution of OCTN variants within the IBD5 locus to pediatric onset Crohn's disease. Am J Gastroenterol, 101(6), 1354-1361.doi: 10.1111/j.1572-0241.2006.00564.x.

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., . . . Daly, M. J. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet, 40*(8), 955-962. doi: 10.1038/ng.175

Baumgart, D. C., & Carding, S. R. (2007). Inflammatory bowel disease: cause and immunobiology. *Lancet, 369*(9573), 1627-1640. doi: 10.1016/s0140-6736(07)60750-8

Bennett, K. M., Liu, J., Hoelting, C., & Stoll, J. (2011).Expression and analysis of two novel rat organic cation transporter homologs, SLC22A17 and SLC22A23.*Mol Cell Biochem, 352*(1-2), 143-154. doi: 10.1007/s11010-011-0748-y.

Bennett, R. A., Rubin, P. H., & Present, D. H. (1991).Frequency of inflammatory bowel disease in offspring of couples both presenting with inflammatory bowel disease.Gastroenterology, 100(6), 1638-1643.

Bernstein, C. N., Blanchard, J. F., Rawsthorne, P., & Wajda, A. (1999). Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study. *Am J Epidemiol, 149*(10), 916-924.

Bernstein, C. N., Wajda, A., Svenson, L. W., MacKenzie, A., Koehoorn, M., Jackson, M., . . . Blanchard, J. F. (2006). The epidemiology of inflammatory bowel disease in Canada: a population-based study. Am J Gastroenterol, 101(7), 1559-1568.doi: 10.1111/j.1572-0241.2006.00603.x.

Binder, V., Both, H., Hansen, P. K., Hendriksen, C., Kreiner, S., & Torp-Pedersen, K. (1982). Incidence and prevalence of ulcerative colitis and Crohn's disease in the County of Copenhagen, 1962 to 1978. Gastroenterology, 83(3), 563-568.

Blanchard, J. F., Bernstein, C. N., Wajda, A., & Rawsthorne, P. (2001). Small-area variations and sociodemographic correlates for the incidence of Crohn's disease and ulcerative colitis. Am J Epidemiol, 154(4), 328-335.

Brant, S. R., Wang, M. H., Rawsthorne, P., Sargent, M., Datta, L. W., Nouvet, F., . . . Bernstein, C. N. (2007). A population-based case-control study of CARD15 and other risk factors in Crohn's disease and ulcerative colitis.*Am J Gastroenterol, 102*(2), 313-323.doi: 10.1111/j.1572-0241.2006.00926.x

Calkins, B. M., & Mendeloff, A. I. (1986).Epidemiology of inflammatory bowel disease.*Epidemiol Rev, 8*, 60-91.

Cedernaes, J., Olszewski, P. K., Almen, M. S., Stephansson, O., Levine, A. S., Fredriksson, R., . . . Schioth, H. B. (2011). Comprehensive analysis of localization of 78 solute carrier genes throughout the subsections of the rat gastrointestinal tract.*Biochem Biophys Res Commun, 411*(4), 702-707. doi: 10.1016/j.bbrc.2011.07.005.

Chen, L. H., Kuo, W. H., Tsai, M. H., Chen, P. C., Hsiao, C. K., Chuang, E. Y., . . . Chang, K. J. (2011). Identification of prognostic genes for recurrent risk prediction in triple negative breast cancer patients in Taiwan.PLoS One, 6(11), e28222.doi: 10.1371/journal.pone.0028222.

Chusainow, J., Yang, Y. S., Yeo, J. H., Toh, P. C., Asvadi, P., Wong, N. S., & Yap, M. G. (2009). A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? Biotechnol Bioeng, 102(4), 1182-1196. doi: 10.1002/bit.22158.

Cucchiara S, Latiano A, Palmieri O, Staiano AM, D'Inca` R, Guariso G, Vieni G, Rutigliano V,Borrelli O, Valvano MR, et al. (2007) Role of CARD15, DLG5 and OCTN genes polymorphismsin children with inflammatory bowel diseases. World J Gastroenterol 13:1221–1229.

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., & Lander, E. S. (2001). High-resolution haplotype structure in the human genome. Nat Genet, 29(2), 229-232. doi: 10.1038/ng1001-229.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., . . .Notredame, C. (2011). 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(Web Server issue), W13-17.doi: 10.1093/nar/gkr245.

Dukes, J. D., Whitley, P., & Chalmers, A. D. (2011). The MDCK variety pack: choosing the right strain. *BMC Cell Biology*, *12*, 43.doi:10.1186/1471-2121-12-43.

Ek, W. E., D'Amato, M., &Halfvarson, J. (2014).The history of genetics in inflammatory bowel disease.Annals of Gastroenterology : Quarterly Publication of the Hellenic Society of Gastroenterology, 27(4), 294–303.

Facts about Inflammatory Bowel Diseases (May 1, 2011). From http://www.ccfa.org/resources/facts-about-inflammatory.html

Fiocchi, C. (1998). Inflammatory bowel disease: etiology and pathogenesis. *Gastroenterology, 115*(1), 182-205.

Fisher, S. A., Hampe, J., Onnie, C. M., Daly, M. J., Curley, C., Purcell, S., . . . Mathew, C. G. (2006). Direct or indirect association in a complex disease: the role of SLC22A4 and SLC22A5 functional variants in Crohn disease. Hum Mutat, 27(8), 778-785. doi: 10.1002/humu.20358.

Flow Cytometry BestProtocols®. (2013). eBioscience (Ed.), (pp. 1-3).Retrieved May 20, 2015, from http://www.ebioscience.com.

Franke, A., Balschun, T., Karlsen, T. H., Hedderich, J., May, S., Lu, T., . . . Schreiber, S. (2008). Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nat Genet, 40*(6), 713-715. doi: 10.1038/ng.148.

Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., . . . Parkes, M. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet, 42*(12), 1118-1125. doi: 10.1038/ng.717.

Garrett, Q., Xu, S., Simmons, P. A., Vehige, J., Flanagan, J. L., & Willcox, M. D. (2008). Expression and localization of carnitine/organic cation transporter OCTN1 and OCTN2 in ocular epithelium.*Invest Ophthalmol Vis Sci, 49*(11), 4844-4849. doi: 10.1167/iovs.07-1528.

Giallourakis, C., Stoll, M., Miller, K., Hampe, J., Lander, E. S., Daly, M. J., . . . Rioux, J. D. (2003). 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(1), 205-211. doi: 10.1086/376417.

Gilchrist, S. E., & Alcorn, J. (2010). Lactation stage-dependent expression of transporters in rat whole mammary gland and primary mammary epithelial organoids. *Fundam Clin Pharmacol, 24*(2), 205-214. doi: 10.1111/j.1472-8206.2009.00760.x.

Grigat, S., Harlfinger, S., Pal, S., Striebinger, R., Golz, S., Geerts, A., . . .Grundemann, D. (2007). Probing the substrate specificity of the ergothioneine transporter with methimazole, hercynine, and organic cations.BiochemPharmacol, 74(2), 309-316. doi: 10.1016/j.bcp.2007.04.015.

Grundemann, D., Harlfinger, S., Golz, S., Geerts, A., Lazar, A., Berkels, R., . . . Schomig, E. (2005). Discovery of the ergothioneine transporter.Proc Natl AcadSci U S A, 102(14), 5256-5261.doi: 10.1073/pnas.0408624102.

Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E., & Jarnerot, G. (2003). Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology, 124*(7), 1767-1773.

Hediger, M. A., Clemencon, B., Burrier, R. E., & Bruford, E. A. (2013). The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol Aspects Med, 34*(2-3), 95-107. doi: 10.1016/j.mam.2012.12.009.

Hediger, M. A., Romero, M. F., Peng, J. B., Rolfs, A., Takanaga, H., &Bruford, E. A. (2004). The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. Pflugers Arch, 447(5), 465-468. doi: 10.1007/s00424-003-1192-y.

Hennerdal, A., & Elofsson, A. (2011).Rapid membrane protein topology prediction.*Bioinformatics, 27*(9), 1322-1323. doi: 10.1093/bioinformatics/btr119.

Herzenberg, L., & Sweet, R. (1976).Fluorescence-activated cell sorting.Sci Am, 234(3), 108-117.

Hidalgo, I. J., Raub, T. J., & Borchardt, R. T. (1989).Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability.*Gastroenterology, 96*(3), 736-749.

Horvath, G., Schmid, N., Fragoso, M. A., Schmid, A., Conner, G. E., Salathe, M., & Wanner, A. (2007). Epithelial organic cation transporters ensure pH-dependent drug absorption in the airway. *Am J Respir Cell Mol Biol, 36*(1), 53-60.doi: 10.1165/rcmb.2006-0230OC.

Hou, J. K., Abraham, B., & El-Serag, H. (2011). Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. *Am J Gastroenterol, 106*(4), 563-573.doi: 10.1038/ajg.2011.44.

Huff, C. D., Witherspoon, D. J., Zhang, Y., Gatenbee, C., Denson, L. A., Kugathasan, S., . . .Guthery, S. L. (2012). Crohn's disease and genetic hitchhiking at IBD5.MolBiolEvol, 29(1), 101-111.doi: 10.1093/molbev/msr151.

Jacobsson, J. A., Haitina, T., Lindblom, J., & Fredriksson, R. (2007).Identification of six putative human transporters with structural similarity to the drug transporter SLC22 family.*Genomics, 90*(5), 595-609. doi: 10.1016/j.ygeno.2007.03.017.

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., … Cho, J. H. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature, 491(7422), 119–124. doi:10.1038/nature11582.

Juyal, G., Prasad, P., Senapati, S., Midha, V., Sood, A., Amre, D., . . . Bk, T. (2011). An investigation of genome-wide studies reported susceptibility loci for ulcerative colitis shows limited replication in north Indians. *PLoS One, 6*(1), e16565. doi: 10.1371/journal.pone.0016565.

Kaplan, G. G., Hubbard, J., Korzenik, J., Sands, B. E., Panaccione, R., Ghosh, S., . . . Villeneuve, P. J. (2010). The inflammatory bowel diseases and ambient air pollution: a novel association. *Am J Gastroenterol, 105*(11), 2412-2419. doi: 10.1038/ajg.2010.252.

Kaser, A., Zeissig, S., & Blumberg, R. S. (2010). Inflammatory bowel disease.*Annu Rev Immunol, 28*, 573-621. doi: 10.1146/annurev-immunol-030409-101225.

Kirsner, J. B., & Spencer, J. A. (1963). Family occurrences of ulcerative colitis, regional enteritis, and ileocolitis. *Ann Intern Med, 59*, 133-144.

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

Koepsell, H. (2013). The SLC22 family with transporters of organic cations, anions and zwitterions.*Mol Aspects Med, 34*(2-3), 413-435. doi: 10.1016/j.mam.2012.10.010.

Koepsell, H., & Endou, H. (2004).The SLC22 drug transporter family.*Pflugers Arch, 447*(5), 666-676. doi: 10.1007/s00424-003-1089-9.

Kurata, J. H., Kantor-Fish, S., Frankl, H., Godby, P., & Vadheim, C. M. (1992). Crohn's disease among ethnic groups in a large health maintenance organization. Gastroenterology, 102(6), 1940-1948.

Kurata, J. H., Kantor-Fish, S., Frankl, H., Godby, P., & Vadheim, C. M. (1992). Crohn's disease among ethnic groups in a large health maintenance organization.*Gastroenterology, 102*(6), 1940-1948.

Laharie, D., Debeugny, S., Peeters, M., Van Gossum, A., Gower-Rousseau, C., Belaiche, J., . . .Colombel, J. F. (2001). Inflammatory bowel disease in spouses and their offspring.Gastroenterology, 120(4), 816-819.

Lapidus, A. (2006). Crohn's disease in Stockholm County during 1990-2001: an epidemiological update. World J Gastroenterol, 12(1), 75-81.

Lashner, B. A. (1995). Epidemiology of inflammatory bowel disease.*Gastroenterol Clin North Am, 24*(3), 467-474.

Lees, C. W., Barrett, J. C., Parkes, M., & Satsangi, J. (2011). New IBD genetics: common pathways with other diseases. *Gut, 60*(12), 1739-1753. doi: 10.1136/gut.2009.199679.

Lin, Z., Nelson, L., Franke, A., Poritz, L., Li, T. Y., Wu, R., . . . Koltun, W. A. (2010). OCTN1 variant L503F is associated with familial and sporadic inflammatory bowel disease. *J Crohns Colitis, 4*(2), 132-138. doi: 10.1016/j.crohns.2009.09.003.

Loftus, C. G., Loftus, E. V., Jr., Harmsen, W. S., Zinsmeister, A. R., Tremaine, W. J., Melton, L. J., 3rd, & Sandborn, W. J. (2007). Update on the incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota, 1940-2000. Inflamm Bowel Dis, 13(3), 254-261. doi: 10.1002/ibd.20029.

Loftus, E. V., Jr. (2004). Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. Gastroenterology, 126(6), 1504-1517.

Manolio, T. A. (2010). Genomewide association studies and assessment of the risk of disease.NEngl J Med, 363(2), 166-176.doi: 10.1056/NEJMra0905980.

Matricon, J., Barnich, N., & Ardid, D. (2010).Immunopathogenesis of inflammatory bowel disease.*Self Nonself, 1*(4), 299-309. doi: 10.4161/self.1.4.13560.

Mattar, M. C., Lough, D., Pishvaian, M. J., & Charabaty, A. (2011).Current Management of Inflammatory Bowel Disease and Colorectal Cancer. Gastrointestinal Cancer Research : GCR, 4(2), 53–61.

McGovern, D. P., Gardet, A., Torkvist, L., Goyette, P., Essers, J., Taylor, K. D., . . . Seielstad, M. (2010). Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet, 42*(4), 332-337. doi: 10.1038/ng.549.

Mikhailova, T. L., Sishkova, E., Poniewierka, E., Zhidkov, K. P., Bakulin, I. G., Kupcinskas, L., Lesniakowski, K., Grinevich, V. B., Malecka-Panas, E., Ardizzone, S., D'Arienzo, A., Valpiani, D., Koch, M., Denapiene, G., Vago, G., Fociani, P., Zerbi, P., Ceracchi, M., Camerini, R. and Gasbarrini, G. (2011), Randomised clinical trial: the efficacy and safety of propionyl-l-carnitine therapy in patients with ulcerative colitis receiving

stable oral treatment. Alimentary Pharmacology & Therapeutics, 34: 1088–1097. doi: 10.1111/j.1365-2036.2011.04844.x

Mirza, M. M., Fisher, S. A., King, K., Cuthbert, A. P., Hampe, J., Sanderson, J., . . . Christopher Mathew, G. (2003). Genetic Evidence for Interaction of the 5q31 Cytokine Locus and the CARD15 Gene in Crohn Disease. The American Journal of Human Genetics, 72(4), 1018-1022. doi: 10.1086/373880.

Moise Pinto, S. R.-L., Marie-Dominique Appay, Michele Kedinger, Nicole Triadou, Elisabeth Dussulx , Brigitte Lacroix, & Patricia Simon-Assmann, K. H., Jorgen Fogh, and Alain Zweibumm. (1983). Enterocyte-like Differentiation and Polarization of the Human Colon Carcinoma Cell Line Caco-2 in Culture.BioI. Cell, 47, 323-330.

Monsen, U., Bernell, O., Johansson, C., & Hellers, G. (1991).Prevalence of inflammatory bowel disease among relatives of patients with Crohn's disease.*Scand J Gastroenterol, 26*(3), 302-306.

Munkholm, P., Langholz, E., Nielsen, O. H., Kreiner, S., & Binder, V. (1992). Incidence and prevalence of Crohn's disease in the county of Copenhagen, 1962-87: a sixfold increase in incidence. Scand J Gastroenterol, 27(7), 609-614.

Negoro, K., McGovern, D. P. B., Kinouchi, Y., Takahashi, S., Lench, N. J., Shimosegawa, T., … van Heel, D. A. (2003). Analysis of the IBD5 locus and potential gene-gene interactions in Crohn's disease. Gut, 52(4), 541–546.

Newman, B., & Siminovitch, K. A. (2005). Recent advances in the genetics of inflammatory bowel disease. *Curr Opin Gastroenterol, 21*(4), 401-407.

Ng, S. C., Bernstein, C. N., Vatn, M. H., Lakatos, P. L., Loftus, E. V., Jr., Tysk, C., . . . Colombel, J. F. (2013). Geographical variability and environmental risk factors in inflammatory bowel disease.Gut, 62(4), 630-649.doi: 10.1136/gutjnl-2012-303661.

Nicolae, A., Wahrheit, J., Bahnemann, J., Zeng, A.-P., & Heinzle, E. (2014). Non-stationary 13C metabolic flux analysis of Chinese hamster ovary cells in batch culture using

extracellular labeling highlights metabolic reversibility and compartmentation. BMC Systems Biology, 8, 50.doi:10.1186/1752-0509-8-50.

Noble, C. L., & Arnott, I. D. (2008). What is the risk that a child will develop inflammatory bowel disease if 1 or both parents have IBD? *Inflamm Bowel Dis, 14 Suppl 2*, S22-23. doi: 10.1002/ibd.20575.

Noble, C. L., Abbas, A. R., Cornelius, J., Lees, C. W., Ho, G. T., Toy, K., . . . Diehl, L. (2008). Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. Gut, 57(10), 1398-1405. doi: 10.1136/gut.2008.148395.

Noble, C. L., Nimmo, E. R., Drummond, H., Ho, G. T., Tenesa, A., Smith, L., . . .Satsangi, J. (2005). The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease.Gastroenterology, 129(6), 1854-1864. doi: 10.1053/j.gastro.2005.09.025.

Orholm, M., Binder, V., Sorensen, T. I., Rasmussen, L. P., & Kyvik, K. O. (2000). Concordance of inflammatory bowel disease among Danish twins.Results of a nationwide study.*Scand J Gastroenterol, 35*(10), 1075-1081.

Palmieri OL, Anna ; Maglietta, Rosalia ; Palumbo, Orazio ; Corritore, Giuseppe ; Latiano, Tiziana ; Martino, Giuseppina ; Carella, Massimo ; Ancona, Nicola ; Bossa, Fabrizio ; Andriulli, Angelo ; Annese, Vito. (2011). Dissection of the Crohn's Disease Transcriptome of 71 Loci Using Genome-Wide *Microarrays.Gastroenterology*,140(5):S-272-S-3.

Pearson, T. A., &Manolio, T. A. (2008). How to interpret a genome-wide association study.Jama, 299(11), 1335-1344.doi: 10.1001/jama.299.11.1335.

Pedersen, J., Coskun, M., Soendergaard, C., Salem, M., & Nielsen, O. H. (2014). Inflammatory pathways of importance for management of inflammatory bowel disease.*World J Gastroenterol, 20*(1), 64-77. doi: 10.3748/wjg.v20.i1.64.

Peltekova, V. D., Wintle, R. F., Rubin, L. A., Amos, C. I., Huang, Q., Gu, X., . . . Siminovitch, K. A. (2004). Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet, 36*(5), 471-475. doi: 10.1038/ng1339.

Pipes, L., Li, S., Bozinoski, M., Palermo, R., Peng, X., Blood, P., . . . Katze, M. G. (2013). The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics.*Nucleic Acids Res, 41*(Database issue), D906-914.doi: 10.1093/nar/gks1268.

Pochini, L., Scalise, M., Galluccio, M., & Indiveri, C. (2013). OCTN cation transporters in health and disease: role as drug targets and assay development. *J Biomol Screen, 18*(8), 851-867. doi: 10.1177/1087057113493006.

Podolsky, D. K. (2002).Inflammatory bowel disease.*N Engl J Med, 347*(6), 417-429. doi: 10.1056/NEJMra020831.

Reinhard, C., & Rioux, J. D. (2006). Role of the IBD5 susceptibility locus in the inflammatory bowel diseases.Inflamm Bowel Dis, 12(3), 227-238. doi: 10.1097/01.MIB.0000202413.16360.a1

Repnik, K., & Potocnik, U. (2011). 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(9), 1081-1091. doi: 10.1007/s00535-011-0426-6.

Rioux, J. D., Silverberg, M. S., Daly, M. J., Steinhart, A. H., McLeod, R. S., Griffiths, A. M., . . . Siminovitch, K. A. (2000). Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet, 66*(6), 1863-1870. doi: 10.1086/302913.

Rocchi, A., Benchimol, E. I., Bernstein, C. N., Bitton, A., Feagan, B., Panaccione, R., . . . Ghosh, S. (2012). Inflammatory bowel disease: a Canadian burden of illness review. *Can J Gastroenterol, 26*(11), 811-817.

Sambuy, Y., De Angelis, I., Ranaldi, G., Scarino, M. L., Stammati, A., &Zucco, F. (2005). The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell BiolToxicol, 21*(1), 1-26. doi: 10.1007/s10565-005-0085-6.

Schlessinger, A., Khuri, N., Giacomini, K. M., & Sali, A. (2013).Molecular modeling and ligand docking for solute carrier (SLC) transporters.*Curr Top Med Chem, 13*(7), 843-856.

Schlessinger, A., Khuri, N., Giacomini, K. M., &Sali, A. (2013).Molecular modeling and ligand docking for Solute Carrier (SLC) transporters.Current Topics in Medicinal Chemistry, 13(7), 843–856.

Sedlack, R. E., Whisnant, J., Elveback, L. R., & Kurland, L. T. (1980).Incidence of Crohn's disease in Olmsted County, Minnesota, 1935-1975. Am J Epidemiol, 112(6), 759-763.

Serrano Leon, A., Amir Shaghaghi, M., Yurkova, N., Bernstein, C. N., El-Gabalawy, H., & Eck, P. (2014). Single-nucleotide polymorphisms in SLC22A23 are associated with ulcerative colitis in a Canadian white cohort. *Am J Clin Nutr, 100*(1), 289-294. doi: 10.3945/ajcn.113.080549.

Shaw, S. Y., Blanchard, J. F., & Bernstein, C. N. (2010). Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease.*Am J Gastroenterol, 105*(12), 2687-2692. doi: 10.1038/ajg.2010.398.

Silverberg MS, Duerr RH, Brant SR, Bromfield G, Datta LW, Jani N, Kane SV, Rotter JI, PhilipSchumm L, Hillary Steinhart A, et al. (2007). Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. Eur J Hum Genet 15:328–335.

Sonnenberg, A., McCarty, D. J., & Jacobsen, S. J. (1991).Geographic variation of inflammatory bowel disease within the United States.*Gastroenterology, 100*(1), 143-149.

Taubert, D., Grimberg, G., Jung, N., Rubbert, A., & Schomig, E. (2005).Functional role of the 503F variant of the organic cation transporter OCTN1 in Crohn's disease.Gut, 54(10), 1505-1506. doi: 10.1136/gut.2005.076083.

Taubert, D., Jung, N., Goeser, T., & Schomig, E. (2009). Increased ergothioneine tissue concentrations in carriers of the Crohn's disease risk-associated 503F variant of the organic cation transporter OCTN1. Gut, 58(2), 312-314. doi: 10.1136/gut.2008.164418.

The Impact of IBD in Canada 2012 Final Report and Recommendations. (2012). Retrieved from www.ccfc.ca, www.fcmii.ca

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

Török, H.-P., Glas, J., Tonenchi, L., Lohse, P., Müller-Myhsok, B., Limbersky, O., … Folwaczny, C. (2005). Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease.Gut, 54(10), 1421–1427. doi:10.1136/gut.2005.066340.

Ventham, N. T., Kennedy, N. A., Nimmo, E. R., & Satsangi, J. (2013). Beyond Gene Discovery in Inflammatory Bowel Disease: The Emerging Role of Epigenetics. Gastroenterology, 145(2), 293–308. doi:10.1053/j.gastro.2013.05.050.

Vermeire, S., Pierik, M., Hlavaty, T., Claessens, G., van Schuerbeeck, N., Joossens, S., . . . Rutgeerts, P. (2005). Association of organic cation transporter risk haplotype with perianal penetrating Crohn's disease but not with susceptibility to IBD. Gastroenterology, 129(6), 1845-1853. doi: 10.1053/j.gastro.2005.10.006.

Viklund, H., & Elofsson, A. (2004). Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci, 13*(7), 1908-1917. doi: 10.1110/ps.04625404.

Waller S, Tremelling M, Bredin F, Godfrey L, Howson J, and Parkes M (2006) Evidence for association of OCTN genes and IBD5 with ulcerative colitis. Gut 55:809–814.

Weersma, R. K., van Dullemen, H. M., van der Steege, G., Nolte, I. M., Kleibeuker, J. H., & Dijkstra, G. (2007). Review article: Inflammatory bowel disease and genetics. Aliment PharmacolTher, 26 Suppl 2, 57-65. doi: 10.1111/j.1365-2036.2007.03476.x.

Yabuuchi, H., Tamai, I., Nezu, J., Sakamoto, K., Oku, A., Shimane, M., . . . Tsuji, A. (1999). Novel membrane transporter OCTN1 mediates multispecific, bidirectional, and pH-dependent transport of organic cations. J PharmacolExpTher, 289(2), 768-773.

Yamazaki, K., Takazoe, M., Tanaka, T., Ichimori, T., Saito, S., Iida, A., . . . Nakamura, Y. (2004). Association analysis of SLC22A4, SLC22A5 and DLG5 in Japanese patients with Crohn disease. J Hum Genet, 49(12), 664-668. doi: 10.1007/s10038-004-0204-x

Zhang, T., Xiang, C.D., Gale, D., Carreiro, S., Wu, E.Y., Zhang, E.Y., 2008b. Drug transporter and cytochrome P450 mRNA expression in human ocular barriers:implications for ocular drug disposition. Drug Metab.Dispos. 36, 1300–1307.

Zhao, Z. Z., Croft, L., Nyholt, D. R., Chapman, B., Treloar, S. A., Hull, M. L., & Montgomery, G. W. (2011). Evaluation of polymorphisms in predicted target sites for micro RNAs differentially expressed in endometriosis. *Mol Hum Reprod, 17*(2), 92-103. doi: 10.1093/molehr/gaq084.