INVOLVEMENT OF GALNT12 VARIANTS IN INHERITED SUSCEPTIBILITY TO COLORECTAL CANCER

ERICA L. CLARKE

INVOLVEMENT OF *GALNT12* VARIANTS IN INHERITED SUSCEPTIBILITY TO COLORECTAL CANCER

by

Erica L. Clarke

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the

requirements for the degree of

Master of Science

Discipline of Genetics/ Faculty of Medicine

Memorial University of Newfoundland

April, 2012

St. John's

Newfoundland

Abstract

Colorectal cancer (CRC) is a major health concern in the Canadian province of Newfoundland and Labrador (NL), as this population has the second highest incidence of CRC in North America. Previous mutation analysis in the NL CRC population has revealed that the high frequency of hereditary CRC in NL is likely attributable to novel susceptibility genes. A recent study suggests that *N-acetylgalactosaminyltransferase 12* (*GALNT12*) may be a novel gene associated with CRC susceptibility. The purpose of this study was to investigate the impact of *GALNT12* germline variants in CRC families from NL and in 481 cases from the Newfoundland Colorectal Cancer Registry (NFCCR) with an unknown molecular basis for their disease.

GALNT12 was first sequenced in 129 CRC families to investigate if variants within this gene segregate with the disease. Probands included in the study have a high family incidence of CRC and were referred to the Provincial Medical Genetics Program. DNA from 481 patients from the NFCCR was also sequenced in exons 4 and 6 to identify causal variants.

Two putatively pathogenic variants were identified in the *GALNT12* gene within four clinical family probands - c.907G>A, p.Asp303Asn (three probands), and c.1187A>G, p.Tyr396Cys. Two families showed partial segregation of c.907G>A with CRC or adenomas. The c.907G>A variant was identified in an additional two cases after sequencing 481 cases from the NFCCR. Another variant, c.889C>T, was also identified that was previously reported in a CRC case. This study provides some support for the

association between *GALNT12* and CRC susceptibility and suggests that further investigations should be performed.

Acknowledgements

First and foremost, I would like to thank Dr. Michael Woods for his patience, understanding and guidance over the last 2 years. Regardless if it was a bad bout of contamination in the lab or an illness of a family member, he was always kind and understanding. Thank you for pushing me to discover the answers to my own questions and also teaching me when I needed to be taught. I would also like to thank my supervisory committee; Dr. Ban Younghusband, Dr. Roger Green and Dr. Jane Green for their contribution and direction over the course of my project. Pursuing this degree has been a great educational experience and has enabled me to grow so much academically.

I would like to especially thank Ms. Krista Mahoney, Mrs. Amanda Dohey and Ms. Michelle Simms for helping me in the lab and giving me advice about my project. Without you guys I would probably still be in the lab trying to collect my results! I would also like to thank Dr. Betty Dicks for her help over the course of my program and Mrs. Deborah Quinlan (my genetics mom) for all her support with Journal Club and so many other things I needed over the last 2 years.

Thank you to the Discipline of Genetics, the Faculty of Medicine Colorectal Cancer Team and the Provincial Medical Genetics Program for their involvement in the study. Thank you to the Atlantic Medical Genetics and Genomics Initiative, Canadian Institute for Health Research, and the Canadian Cancer Society for funding the study. Most importantly, thank you to those who participated in the study. Without your participation this project would not be possible.

Also a big thanks to Mr. Fady Kamel, Mr. Ashar Pirzada and Dr. Jennifer Shea for the great friendships and the many, much needed coffee breaks over the last 2 years. Especially thanks to you Dr. Jen for the endless advice, guidance and many needed chats throughout my studies. I could not have asked for a better mentor and friend. I would like to thank my family and friends for their support over the last two years - particularly to Wes Whalen, for everything, thank you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

8-OxoG	8-hydroxyguanine
μL	microliter
μM	micromole
AC-I	Amsterdam Criteria I
AC-II	Amsterdam Criteria II
AFAP	Attenuated Familial Adenomatous Polyposis
APC	*Adenomatous Polyposis Coli*
BMPR1A	*Bone Morphogenetic Protein Receptor, type IA*
BRAF	*Serine/Threonine protein kinase B-Raf*
BUB1B	*Budding Uninhibited by Benzimidazoles 1 homolog beta*
BUBR1	Budding Uninhibited by Benzimidazoles 1 homolog beta protein
Ca	Cancer
cDNA	Complimentary DNA
cM	Centimorgan
CRC	Colorectal Cancer
DNA	Deoxyribonucleic Acid
dNTP	Dexoyribonucleotide Triphosphate
EDTA	Ethylenediaminetetraacetic Acid
ExoI	Exo Nuclease I
FAP	Familial Adenomatous Polyposis
FCCTX	Familial Colorectal Cancer Type X
g	grams
GalNAC	*N*-acetylgalactosamine

GALNT12	*N-acetylgalactosaminyltransferase type 12*
HLOD	Heterogeneity LOD Score
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HP	Hyperplastic Polyp
ICG-HNPCC	International Collaborative Group on HNPCC
IHC	Immunohistochemistry
JPS	Juvenile Polyposis Syndrome
KRAS	*Kirsten rat sarcoma-2 viral (v-Ki-ras2) oncogene homolog*
LKB1	*Liver Kinase B1*
LOD	Log of Odds score
LS	Lynch Syndrome
ng	nanogram
MAP	*MUTYH*-associated Polyposis
mg	milligram
MGMT	*O-6-methylguanine-DNA methyltransferase*
min	minutes
mL	milliliter
MLPA	Multiplex Ligation-dependant Probe Amplification
MLH1	*MutL Homologue 1*
MLH3	*MutL Homologue 3*
mM	millimole
MMR	Mismatch Repair
MSH2	*MutS Homologue 2*
MSH6	*MutS Homologue 6*
MSI	Microsatellite Instability

MSS	Microsatellite Stable
MUC	Mucin
MUC1	*Mucin 1; transmembrane*
MUTYH	*MutY Homologue*
NFCCR	Newfoundland Colorectal Cancer Registry
NL	Newfoundland and Labrador
NPL	Non-Parametric Linkage
PCR	Polymerase Chain Reaction
PID	Personal Identification Number
PMGP	Provincial Medical Genetics Program
PMS2	*Postmeiotic Segregation increased 2*
ppGalNac-Ts	*N*-acetylgalactosaminyltransferases
rpm	rotations per minute
s	seconds
SAP	Shrimp Alkaline Phosphatase
SIFT	Sorting Intolerant From Tolerant
SISA	Simplified Segregation Analysis
SMAD4	*SMAD family member 4*
SNP	Single Nucleotide Polymorphism
STK11	*Serine/threonine Kinase 11*
U	Units
UCV	Unclassified Variant
V	volts
WT	Wild type

Chapter 1

1.0 Colorectal Cancer Literature Review

1.1 Colorectal Cancer: An important health risk

Colorectal cancer (CRC) is the third most common form of cancer in developed countries and is diagnosed in approximately 22,550 Canadians annually (Canadian Cancer Society, 2011). CRC is the second leading cause of cancer-related death in Canada, killing an estimated 5,000 men and 4,100 women each year (Canadian Cancer Statistics, 2011). The lifetime risk of developing CRC is approximately 6-7% (Canadian Cancer Statistics, 2011).

Approximately 20% of all colorectal cancer cases can be attributed to an inherited predisposition (Kemp et al, 2004). There are several known forms of hereditary CRC which contribute to 2-6% of all CRC (Kemp et al, 2004); the most common include Lynch syndrome (LS), familial adenomatous polyposis (FAP) and *MUTYH*-associated polyposis (MAP).

1.2 Hereditary Colorectal Cancer

1.2.1 Lynch Syndrome

Lynch Syndrome (LS) is an autosomal dominant genetic predisposition to colon and extra-colonic cancers caused by pathogenic variants in the mismatch repair (MMR) genes (Bellinzzi and Frankel, 2009). LS is the most common form of hereditary colorectal cancer, accounting for 2-5% of all CRC (Lynch, et al., 2009).

In 1966, Henry Lynch and his team of researchers investigated two large families that had a strong family history of CRC which lacked multiple colonic polyps. It was also noted that a variety of extra-colonic cancers were observed in these families. This syndrome was originally named "cancer family syndrome" and was later referred to as hereditary non-polyposis colorectal cancer (HNPCC) (Lynch et al., 2009). Hereditary CRC that is associated with pathogenic variants in the MMR genes is now referred to as Lynch Syndrome. HNPCC includes LS but also refers to a syndrome with a similar phenotype of unknown genetic etiology (Lynch et al, 2009). While CRC is the most frequent cancer observed in LS, there is also an increased risk of developing multiple extra-colonic cancers including endometrial, ovarian, small bowel, stomach, pancreatic, kidney, ureter, hepato-biliary, and brain cancers (Lynch et al, 2003).

Pathogenic variants within the DNA MMR genes (*MSH2, MLH1, MLH3, MSH6, PMS2)* are the genetic changes that cause LS. The MMR process protects cells against DNA errors that arise during replication and recombination (Hsieh & Yamane, 2008). Mutations during replication occur at a rate of 10^{-10} to 10^{-9} nucleotides per cell division (Iyer et al., 2006). The MMR genes code for protein products which are responsible for the post-replicative identification and repair of mismatched nucleotides or small insertions/deletions within DNA. This process is highly conserved from prokaryotes to eukaryotes (Iyer et al., 2006). Individuals with only a single functional copy of a particular MMR gene are vulnerable to insufficient MMR activity if there is a somatic mutation of the second copy within a cell. Inactivation of MMR proteins involved in the MMR process significantly increases the number of spontaneous mutations occurring in

the cell, increasing the risk of developing cancer in the colon and other extra-colonic tissues associated with LS (Kolodner, 1996).

To date there are five MMR genes identified in which variations have been shown to cause LS: *MSH2, MLH1, MLH3, MSH6,* and *PMS2. MLH1* and *MSH2,* account for 80% of the known LS pathogenic variants (Woods et al, 2007). Jenkins and colleagues (2006) investigated the cancer risk for individuals harbouring MMR pathogenic variants. They determined that the cumulative risk for developing CRC by age 70 was 45% (range 29%-62%) for men and 38% (range 19%-51%) for women.

Risk criteria for classifying families at high risk for LS were developed in 1991 by the International Collaborative Group on HNPCC (ICG-HNPCC). As the meeting was held in Amsterdam, Holland, the criteria became known as "Amsterdam Criteria" (AC-I, see appendix A). At this time, the MMR genes associated with LS had not been identified; therefore, AC-I was used to help select families ideal for linkage studies. A modified set of criteria were developed by the same collaborative group in 1999, which were more sensitive to the presence of extra-colonic cancers observed in LS (AC-II, see appendix A). Additional measures were developed to identify individuals at risk for LS. The Bethesda Criteria (Rodrigues-Bias et al, 1997) and the Revised Bethesda Criteria (Umar et al, 2004) were developed to identify individuals who should be screened for the classical molecular markers of LS including microsatellite instability and MMR protein deficiencies in somatic tumour tissue. See appendix B for the complete Revised Bethesda Criteria.

1.2.2 Familial Adenomatous Polyposis and Attenuated Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP) is a colon cancer syndrome that is characterized by the presence of hundreds to thousands of colonic polyps by the second or third decade of life (Nieuwenhuis & Vasen, 2007). FAP is caused by deleterious germline variants in the adenomatous polyposis coli (*APC*) gene and is inherited in an autosomal dominant fashion (Nieuwenhuis & Vasen, 2007). About 1 in 5,000-10,000 individuals are affected by FAP (Kinzler & Vogelstein, 2002). FAP accounts for < 1% of all CRC (Bisgaard et al, 1994).

A clinical diagnosis of FAP is given when more than 100 colonic polyps are identified, however, the actual number of polyps may be greater than 1000 (Bussey, 1975). Due to the large number of polyps developing at a relatively young age in FAP patients, the onset of CRC by the age of 40-50 years is almost inevitable (Bussey, 1975). Sigmoidoscopic screening is recommended by age 10 - 12 years for individuals harbouring *APC* pathogenic variants and a preventative colectomy is usually recommended by 20 years of age (Vasen, 2000).

The clinical phenotype associated with FAP is not limited to CRC. It also includes dental abnormalities, congenital hypertrophy of the retinal pigment epithelium, lipomas, epidermoid cysts, osteomas, upper gastrointestinal polyps and desmoid tumours (Bussey, 1975). FAP patients are also at an increased risk for developing cancers of the duodenum, brain (medulloblastoma), thyroid and hepatobiliary tract (Bussey, 1975).

The chromosomal instability pathway underlies the development of FAP caused by pathogenic variants in the *APC* gene (Miyoshi et al., 1992). This gene is located on

chromosome 5q21-22 and consists of 15 exons (Kinzler et al, 1991). *APC* is a tumour suppressor gene and is involved in many cellular processes such as cell migration, differentiation, transcription and apoptosis (Goss & Groden, 2000; Sieber et al, 2000). FAP patients have one germline causal variant in *APC* within every cell. Some cells can acquire a second somatic mutation or loss of the other allele. A tumour develops from these cells that have acquired a second somatic insult of the *APC* allele. This follows Alfred Knudson's "two-hit hypothesis" of cancer developing from cells that have accumulated DNA damage (Knudson, 1971).

Pathogenic variants occurring within the 5', 3' or exon 9 regions of the *APC* gene are known to cause a less severe phenotype compared to FAP, called Attenuated Familial Adenomatous Polyposis (AFAP) (Sieber et al, 2006). Pathogenic variants associated with AFAP do not affect the APC protein function as severely as FAP pathogenic variants, thus explaining the attenuated phenotype. Patients with AFAP may have fewer polyps (10s-100s) than those with classic FAP and also may have a later onset of CRC; however, the phenotype can be quite variable (Sieber et al, 2006).

1.2.3 MUTYH-associated polyposis: The MUTYH gene

MUTYH-associated polyposis (MAP) is an autosomal recessive condition associated with deleterious variants in the *MUTYH* gene (Al-Tassen et al, 2002). MAP is associated with an increased risk for the development of CRC (Cheadle and Sampson, 2007) and the clinical phenotype of MAP overlaps that of AFAP (10-100 adenomas) and FAP (>100 adenomas).

MAP was recently recognized in 2002 by Al-Tassen and colleagues and its prevalence, natural history and full phenotype are still being defined (Sampson & Jones, 2009). Previous studies that conducted genetic testing for common deleterious variants underlying MAP (p.Tyr179Cys and p.Gly396Asp) indicate that approximately 1% of all CRC cases from Western Europe and North America are associated with this condition (Croitoru et al, 2004; Fleischmann et al, 2004; Wang et al, 2004; Farrington et al, 2005).

Biallelic pathogenic variants in the *MUTYH* gene are responsible for the development of MAP. The *MUTYH* gene produces a protein involved in the oxidative damage repair pathway. MUTYH is a base-excision repair DNA glycosylase that forms a complex with MSH2/MSH6 which functions to help protect cells against the negative mutagenic effects from oxidative damage (Cheadle & Sampson, 2007). Oxidative damage from reactive oxygen species within the cell can produce 8-hydroxyquanine (8-OxoG) from guanine, one of the most stable products formed from oxidative damage (Cheadle & Sampson, 2007). DNA glycosylases are important for excising adenines improperly paired with 8-OxoG instead of cytosine. Failure to correct this error within the DNA, due to pathogenic variants in the *MUTYH* gene producing a non-functional protein, are thought to lead to the hallmark characteristic of the *APC* G:C→T:A pathogenic variants occurring within MAP tumours (Ames & Gold, 1991).

1.2.4 Familial Colorectal Cancer Type X (FCCTX)

Familial colorectal cancer type X (FCCTX) is a term used to describe families that meet the Amsterdam I criteria (AC-I) yet do not have signs of a MMR defect (see Lynch syndrome above). The AC-I and AC-II were created in 1991 and 1999

respectfully, to help identify families with hereditary nonpolyposis colorectal cancer (HNPCC); HNPCC was earlier described by Dr. Henry Lynch as an autosomal dominant syndrome, associated with an increased risk for CRC with an early age of onset (20 years earlier than non-familial CRC) and an increased probability of cancer on the right side compared to sporadic cases. Extra-colonic cancers are also associated with HNPCC including cancers of the stomach, endometrium, small intestine, kidney, ureter, ovary and hepatobiliary tract (Lynch et al., 2009).

The underlying genetic cause of HNPCC was unknown at the time the AC-I and AC-II were created; however, the application of these criteria was instrumental in identifying and characterizing families for studies which lead to the identification of germline mutations in the MMR genes. "Lynch syndrome" is now used to describe families that harbour MMR gene mutations and the term "Familial Colorectal Cancer Type X" was then assigned to those families that fulfilled AC-I that did not have a DNA mismatch-repair defect identified (Lindor, 2009). The "X" was used to illustrate the unknown etiology of the condition, genetic or otherwise. There has been a call to retire the term HNPCC, since the identification of the involvement of the MMR genes, as it has been improperly used to describe Lynch syndrome (Jass, 2006).

According to Lindor (2009), "FCCTX is a heterogeneous grouping of: (1) random congregations of a common tumour within a family; (2) congregations of a tumour related to shared lifestyle within a family; (3) polygenic predisposition of a family; (4) some undefined monogenic conditions."

Several studies on FCCTX have furthered knowledge about the characteristics of this disorder. It has been shown that there is an older age of onset of CRC in FCCTX cases compared to that of LS cases (55 years verses 41 years) (Mueller-Koch et al., 2005; Dove-Edwin et al., 2006; Valle et al., 2007). FCCTX tumours are more often left sided, compared to LS cases, and FCCTX cases have a greater adenoma/carcinoma ratio (Mueller-Koch et al., 2005). A study of 64 Spanish families that fulfilled AC-I confirmed a later age of onset in FCCTX patients and additionally noted that FCCTX patients were more likely to have left-sided, mucinous tumours and less likely to have multiple primary tumours compared to LS (Valle et al., 2007). It is not known whether FCCTX patients are at risk for the development of extra-colonic cancers. It is important to note that the group of families categorized as having FCCTX are likely a genetically heterogeneous group, with various genetic entities responsible for the hereditary CRC in the different families (M.O. Woods, personal communication, May 2011).

1.2.5 Other Colorectal Cancer Syndromes and Associated Genes

Other than the more common hereditary CRC syndromes, there are other rare cancer syndromes that are associated with known genes contributing to hereditary CRC. Turcot's syndrome is a clinical diagnosis defined by brain tumours as well as polyposis and/or CRC. This includes families with medulloblastoma and *APC* mutations, and also families with MMR mutations where the brain tumour is a glioma or astrocytoma (Hamilton et al., 1995).

Muir-Torre syndrome is a particular subgroup of LS where sebaceous cancers of the skin occur in addition to the other manifestations of LS. Similar to other LS tumours, sebaceous cancers of the skin are also characterized by microsatellite instability (Entius et

al., 2000). Pathogenic variants in the MMR genes *MSH2* and *MLH1* are responsible for the occurrence of this condition. Muir-Torre Syndrome follows an autosomal dominant mode of inheritance, is highly penetrant and has variable expression (Prieto, 2011).

Constitutive MMR deficiency is a separate sub-type of CRC and is caused by biallelic pathogenic variants in the MMR genes. Individuals who inherit homozygous or compound heterozygous mutations develop cancer within the first 10 years of life (Felton et al., 2007). The age of onset for these malignancies generally depends on the severity of the MMR mutation. Pathogenic variants that produce an MMR protein with a complete loss of function usually present with brain and hematological cancers in early childhood (Felton et al., 2007). Individuals that inherit biallelic missense mutations with residual function may present with gastrointestinal cancers or neurofibromas of the skin and tongue within the first decade of life (Felton et al., 2007). Café-au-lait spots are usually present in all individuals that possess biallelic pathogenic variants in any of the MMR genes regardless of the level of residual MMR function (Felton et al., 2007).

Peutz-Jeghers syndrome is characterized by hamartomatous polyps in the gastrointestinal tract and mucocutanous melanin pigmentation, particularly of the lips and oral mucosa. It follows an autosomal dominant pattern of inheritance, incomplete penetrance and variable expression. Individuals with Peutz-Jeghers syndrome are at an increased risk of developing gastrointestinal and non-gastrointestinal tumours (Giardiello & Trimbath, 2006). *STK11* (also known as *LKB1*) has been implicated in 30-70% of sporadic cases of Peutz-Jeghers syndrome and has been responsible for up to 70% of cases with a family history of this condition (Amos et al., 2004). Screening programs for those at risk have now been implemented through genetic testing of the *STK11* gene.

Ongoing studies have attempted to identify the additional genetic entities underlying a significant proportion of unsolved hereditary CRC families. These studies have revealed several genes associated with the disease. A study by Frio and colleagues (2010) found homozygous *BUB1B* splice-site variants in an individual diagnosed with an adenocarcinoma of the ampulla of Vater at 34 years of age, adenomatous polyps 20 years later, and by multiple primary invasive adenocarcinomas within the colon and the stomach. This homozygous variant created a *de novo* splice site affecting the proper transcription of the gene. BUBR1, encoded by *BUB1B*, is important for regulation of the spindle assembly checkpoint during mitosis. It has also been noted that APC and BUBR1 both work together to regulate mitotic cell division (Kaplan et al., 2001). Biallelic pathogenic variants of the *BUB1B* gene had been initially associated with mosaic variegated aneuploidy syndrome and the chromosomal instability pathway. Individuals with mosaic aneuploidy syndrome have growth deficiencies, microcephaly, mental retardation, abnormalities of the central nervous systems and physical dismorphic features in addition to cancer (Garcia-Castillo et al., 2008). The case study presented by Frio and colleagues (2010) identified a specific homozygous change affecting the splice site and the proper transcription of the *BUB1B* gene. This homozygous change was associated with a phenotype of adenomatous polyps and multiple adenocarcinomas within the colon and stomach. This phenotype is hypomorphic compared to those mentioned above, which are known to be associated with other biallelic pathogenic variants in the *BUB1B* gene.

Recently, pathogenic variants in *BMPR1A* have been found in some families that were classified as FCCTX (Nieminen et al, 2011). After genome-wide linkage analysis

was conducted on a family with high *a priori* informativeness, a maximal LOD score of 1.6 was determined for marker *D10S1686* which is mapped to chromosome 10q23. Sequencing of *BMPR1A* revealed a germline tri-nucleotide deletion (c.264-266del, p.Glu88del). This deletion segregated with the adenoma/carcinoma phenotype within a family. Further sequencing in other FCCTX families revealed a proband with a 24-base deletion within intron 1 affecting the 5' portion of exon 2 leading to a predicted skipping of exon 2 and a subsequent frameshift (p.Gly23GlufsX10), which was confirmed by cDNA analysis. *BMPR1A* encodes a type 1 bone morphogenetic protein receptor that is important for activation of SMAD transcriptional regulators. *BMPR1A* has been previously implicated in juvenile polyposis syndrome.

Juvenile polyposis syndrome (JPS) is characterized by the presence of multiple polyps in the gastrointestinal tract including the stomach, small intestine, colon and rectum at a young age, typically by age 20 (Haidle & Howe, 2003). Most of the polyps associated with JPS are benign; however, malignant developments can occur, increasing the lifetime risk of CRC up to 50% (Howe et al., 1998b). This syndrome has an autosomal dominant mode of inheritance and the genes *BMPR1A* and *SMAD4* have been implicated with this disease. About 20% of individuals with this syndrome have a causal variant in *BMPR1A* and an additional 20% have a *SMAD4* deleterious change (Haidle & Howe, 2003). These genes are investigated in molecular genetic testing for clinical and genetic counseling purposes in the management of this syndrome.

1.3 Dietary and Environmental Factors Associated with Colorectal Cancer

Historically, it has been thought that poor (high-fat, low-fiber) diet has been associated with the development of CRC, especially if the diet is coupled with a

sedentary lifestyle, excess caloric intake and other unhealthy habits such as smoking and significant alcohol consumption (Doll et al., 1981; Giovannucci et al., 1992). While some studies have shown that high fat–low fiber diets are linked to an increased risk of developing colorectal adenomas and carcinomas (Giovannucci et al., 1993; Willett et al., 1990), a review of the literature from January 1966 to December 2006 has indicated that many of the dietary influences on the development of CRC are not as well defined as once thought (Ryan-Harshman & Aldoori, 2007).

Studies that investigated the link between red meat and fiber consumption have yielded variable results; however, studies evaluating the link between CRC and the intake of vitamin D, folic acid and calcium have been more reproducible (Ryan-Harshman & Aldoori, 2007).

Minerals such as folic acid, calcium and vitamin D may be an important role in the prevention of colorectal adenomas (Giovannucci et al., 1993; Grau et al., 2003). Furthermore, it has been shown that calcium and vitamin D may work together to reduce recurrence of adenomas. Grau et al. (2003) found that dietary calcium levels were not associated with the presence of adenomas when vitamin D levels were at or below average (29.1 ng/mL) and that a reduced risk of adenomas was only found among individuals that were receiving calcium supplements.

Studies have shown that the regular use of aspirin is associated with a moderate reduction in the risk of developing colonic polyps and CRC (Sandler et al., 2003; Baron et al., 2003; Benamouzig et al., 2003; Logan et al., 2008). Burn and colleagues (2008) investigated the effect of aspirin use on individuals with LS. They found that the use of

aspirin for up to 4 years had no effect on the development of colorectal adenomas or carcinomas. An additional study was conducted by Rothwell et al. (2010) to determine the long-term effectiveness of low aspirin doses on CRC incidence. They found that aspirin, when taken for 5 years or longer, reduced the long-term incidence and associated mortality of CRC. Specifically, taking aspirin prevented cancer in the proximal colon. A more recent study by Burn and colleagues (2011) found that individuals that had a genetic predisposition to LS who took 600 mg of aspirin a day for a mean period of 25 months had a significantly reduced incidence of cancer compared to LS carriers that did not take any aspirin (placebo).

Overall, evidence has shown that diet has a moderate influence on the development of CRC (Ryan-Harshman & Aldoori, 2007). It has been shown that unhealthy lifestyles, such as smoking, low levels of physical exercise and excessive alcohol consumption are associated with ill health including CRC (Chen et al., 1999). A review of the literature indicates a healthy diet that includes vitamin D, calcium and folate may decrease the risk of developing CRC (Ryan-Harshman & Aldoori, 2007). It is very likely that an interaction between genetic factors and environment plays a significant role in the onset of colorectal adenomas and carcinomas (Le Marchand, 2002; Chen et al., 1999).

1.4 Colorectal Polyps and the Associated Risk of Colorectal Cancer

1.4.1 Colorectal Polyp Classifications

A colorectal polyp is defined as a fleshy growth or protrusion of cells that occurs on the epithelial lining of the colon or rectum. Colorectal polyps can be broadly categorized into three histological groups, neoplastic (adenomatous), non-neoplastic and

sub-mucosal (Table 1.1) (Markowitz & Winawer, 1997). Neoplastic (adenomatous) polyps are either pedunculated, with a stalk-like attachment to the wall of the colon or rectum, or sessile, attached to the colon or rectum by a broad or flat base (Figure 1.1) (Markowitz & Winawer, 1997). Identifying neoplastic polyps is important because of their potential to develop into a malignant form. Although it is generally accepted that most CRCs develop from a pre-existing adenomatous polyp, colorectal adenomas occur in up to 50% the general population. Men have a higher frequency of adenomas compared to women and the risk of developing these lesions increases with age (Rickert et al., 1979). Adenomas have been found to occur more frequently on the left side of the colon, sharing the same anatomic distribution as CRC. Adenomas can be further categorized as tubular, tubulovillious or villous. Tubular adenomas account for up to approximately 80% of all adenomas (O'Brien et al., 1990). All adenomas have some degree of dysplasia; mild, moderate or severe (high-grade). The extent of dysplasia is based on glandular architectural distortion and cytological epithelial atypia (Markowitz & Winawer, 1997). The National Polyp Study determined that approximately 86% of adenomas show mild dysplasia, 8% show moderate dysplasia and 6% show severe dysplasia (or carcinoma *in situ*). It was also noted that age, large polyp size and polyps with a villous structure were associated with increased risk for severe dysplasia (O'Brien et al., 1990). Polyps can also be malignant, meaning that they contain malignant cells which extend though the muscularis mucosa into the submucosa. Malignant polyps have the potential to metastasize and occur in approximately 2% of all adenomas (O'Brien et al., 1990). It has also been noted that sessile polyps have a higher potential to metastasize

Table1.1 Histological Categorization of Colorectal Polyps

Neoplastic	Non-Neoplastic	Submucosal
Benign	Hyperplastic	Lymphoid
Mild Dysplasia	Mucosal	Lipoma
Moderate Dysplasia	Inflammatory	Other
High Grade Dysplasia	Hamartomatous	
Severe Dysplasia	Juvenile	
Carcinoma in situ	Peutz-Jeghers	
Malignant		
Invasive Carcinoma		

Adapted from Markowitz & Winawer (1997).

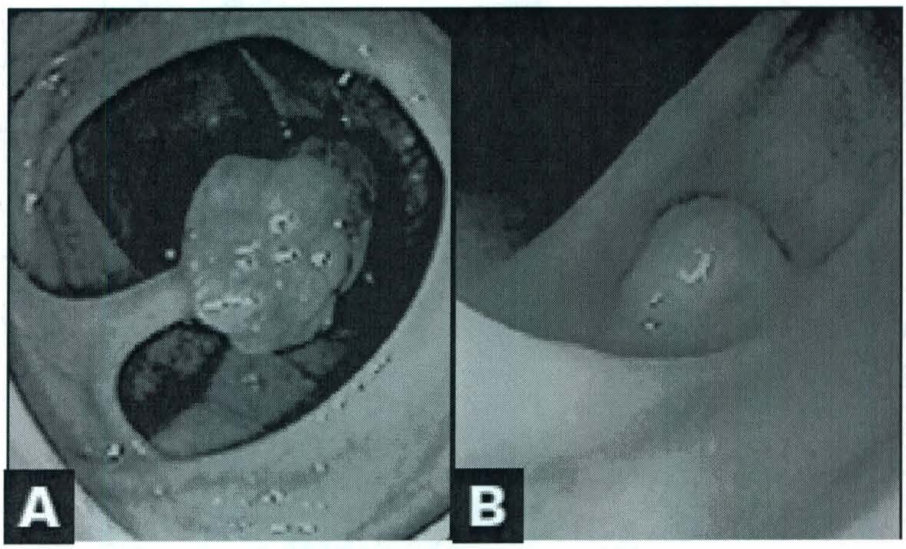

Figure 1.1. Colorectal Polyps. a) pedunculated polyp b) sessile polyp. (Markowitz & Winawer, 1997).

than pedunculated polyps, and are therefore treated more aggressively (Markowitz & Winawer, 1997). Non-neoplastic polyps include hyperplastic, mucosal, inflammatory and harmartomatous polyps. Hyperplastic polyps (HP) are the most frequent form of non-neoplastic polyps, occurring in 20-30% of the general population (Vatn et al., 1982; DiSario et al., 1991). Hyperplastic polyposis is a condition defined as having >20 HPs (John Hopkins Colorectal Cancer, retrieved online). HPs were once thought to carry no malignant potential (Markowitz & Winawer, 1997); however, Torlakovic and colleagues (1996) found that individuals diagnosed with hyperplastic polyposis were at an increased risk for the development of CRC. This same study also reported significant morphological differences between the polyps associated with hyperplastic polyposis and small HPs found in the general population. Hyman and colleagues (2004) found that 7 of 13 patients (54%) diagnosed with at least 30 HPs had developed CRC during the 5 years after HP diagnosis. Recent studies have changed the impression that hyperplastic polyps have no malignant potential and a new serrated neoplasia pathway has been proposed based on the genetic and morphologic distinction of HPs (Jass, 2001).

Serrated precursor lesions have been classified into two types, the hyperplastic polyp and the advanced serrated polyp. The advanced serrated polyp type is further classified into three types: traditional serrated adenomas, sessile serrated adenomas and mixed polyps (Snover et al., 2005). Approximately 80-90% of all serrated polyps are hyperplastic. The advanced serrated polyps are more likely to carry malignant potential compared to HPs. Traditional serrated polyps are a form of adenomatous polyps that are characterized by their serrated edge. These polyps are commonly found in the distal colon and are associated with methylation of the *MGMT* gene as well as pathogenic

variants in the *BRAF* or *KRAS* gene (Jass, 2001). Serrated sessile polyps are distinct from traditional serrated polyps as these lesions are not pedunculated. The sessile serrated adenomas can have T or L-shaped crypts in their architecture and have serration extended to the base of these crypts. The mixed polyp is architecturally a combination of the HP and the traditional serrated adenoma (Jass, 2001).

Mucosal polyps are benign outgrowths of the mucosa that are elevated by the underlying sub-mucosa. These polyps are small in size (0.5cm) and can account for approximately 18% of all resected small polyps (Waye et al., 1988). These polyps do not pose any risk for developing CRC and follow up screening is not needed (Markowitz & Winawer, 1997).

Inflammatory polyps occur within the colon or rectum as a result of severe inflammatory bowel disease. Regenerative repair or ulceration of the rectum or bowel causes these polyps to form. Some inflammatory polyps are formed by residual colonic mucosal tissue that projects into the lumen due to severe ulceration while others are formed from regenerating mucosa and granulation tissue. These types of polyps are usually associated with inflammation due to Crohn's colitis or other conditions causing underlying inflammation of the colon. They can cause abdominal pain or obstruction (Markowitz & Winawer, 1997). Inflammatory bowel disease is associated with an increased risk of CRC, specifically in chronic ulcerative colitis. Patients with Crohn's disease are also at a higher risk for developing CRC (Freeman, 2008).

Hamartomatous polyps contain a mixture of normal tissues within the polyp and have been identified in juvenile polyposis and Peutz-Jeghers syndrome (Markowitz &

Winawer, 1997). Juvenile polyps are generally benign and do not pose a significant malignant risk; however, in rare cases these polyps may show adenomatous or malignant characteristics (Friedman et al., 1982; Tung-Hua et al., 1978).

Sub-mucosal polyps can be classified as lymphoid polyps. Lymphoid polyps are commonly found in the rectum and are caused by the expansion of the lamina propria associated with lymphocytic infiltration. These polyps are benign and are usually asymptomatic occurring in individuals between the ages of 20-60 years (Markowitz & Winawer, 1997). Lipomas are usually found in the proximal colon incidentally during colonoscopy. These asymptomatic, benign entities do not need to be removed and require no follow-up surveillance. (Markowitz & Winawer, 1997).

Efforts to prevent the onset of CRC now focus on screening individuals for colorectal polyps that may develop into a malignant form. There is reasonable data that show that CRC develops from normal mucosa, which then develops into benign adenomas, and subsequently into carcinomas. There is also a great amount of evidence that indicates that the resection of adenomatous polyps prevents CRC (Bond, 2000). Colorectal polyps can be diagnosed through barium radiography, endoscopic procedures or virtual colonoscopy. Most polyps are removed from the colon during colonoscopy by electrocautery techniques. Small polyps (>1 cm) should be removed and biopsied, to indicate if follow-up is needed. Large sessile polyps (<2 cm) should be biopsied and a follow-up colonoscopy should be completed in 3-6 months to confirm a complete resection (Bond, 2000). Colonoscopy screening should be considered for the first degree relatives of individuals that have had adenomatous polyps before the age of 60 years.

CRC is one of the most curable cancers if detected in its early stages. To date, the best way to manage CRC is through prevention. It is recommended that a fecal occult blood test or fecal immunochemical test be conducted every two years for individuals over the age of 50 (Canadian Cancer Society, 2011). Positive fecal tests should be followed by a colonoscopy or double contrast barium enema and flexible sigmoidoscopy for a definitive diagnosis. While screening procedures have been shown to benefit the entire population (Bond, 2000), such procedures are most beneficial to those at high risk based on a family history of the disease and/or those with a known hereditary predisposition.

1.5 Colorectal Cancer and GALNT12

1.5.1 The Association between CRC and Chromosome 9q22

Several studies have identified a link between CRC risk and chromosome 9q22. In 2003, Wiesner and colleagues genotyped 53 familial colorectal neoplasia kindreds and conducted multipoint linkage analysis to test for genetic linkage. The kindreds included an affected sibling pair both diagnosed with either CRC or an adenoma with high grade dysplasia, greater than 1 cm in size, before the age of 65 years. Individuals having FAP or LS were excluded from this study. Linkage was identified between the disease phenotype and the chromosomal region 9q22.2-31.2 ($p= 0.00045$), which followed an autosomal dominant mode of inheritance (Figure 1.2). A study by Skoglund et al. (2006) also identified the chromosomal region 9q22.32-31.1 as a susceptibility locus for CRC and adenomatous polyps. This study was a linkage analysis investigating a large Swedish family who had an identified *MSH2* mutation segregating through the family members;

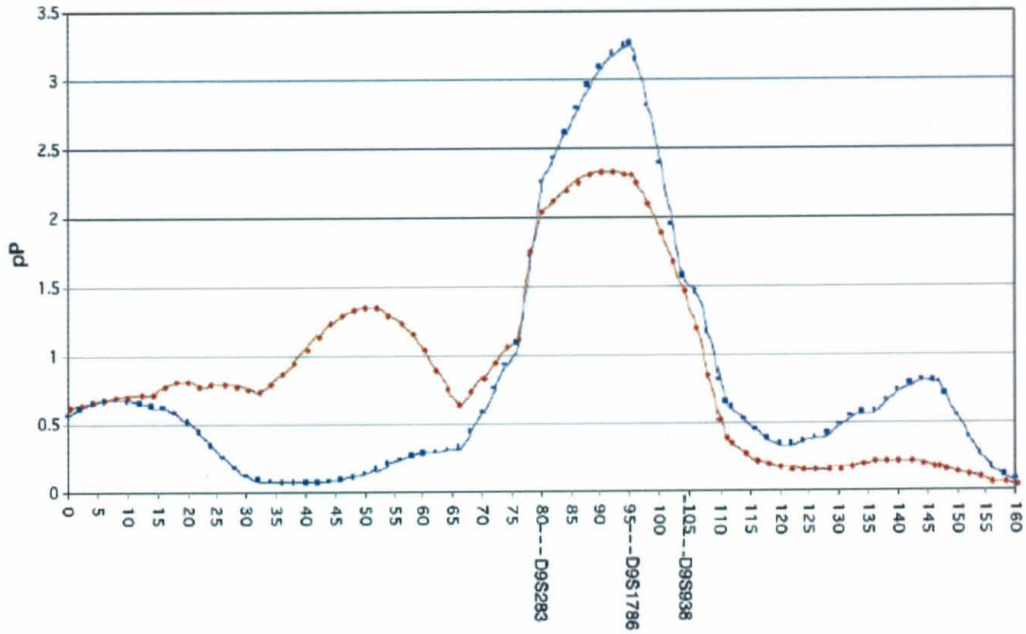

Figure 1.2. Linkage of chromosome 9q22.2-31.2 to CRC phenotype. The x-axis depicts chromosome 9 in cM and the y axis plots pP=-$\log_{10}$(P value for linkage). Red line represents the linkage for concordant twin pairs and the blue line represents linkage using all affected and unaffected siblings. Adapted from Wiesner et al. (2003).

however, this mutation was not present in eight individuals who had been either diagnosed with CRC or adenomatous polyps. The analysis revealed linkage of CRC risk to a region on chromosome 9 with a LOD score of 2.4. These findings not only supported those by Weisner et al. (2003) but also narrowed the region of 9q22.2-31.2 to 9q22.32-31.1, limiting it to 7.9 cM.

An additional study was also conducted in 2006 by Kemp and colleagues who genotyped 16 markers within the suggested susceptibility region of chromosome 9q22.2-31.3, as defined by Weiser et al. (2003). Investigating 57 colorectal cancer families from the United Kingdom, excluding those with known Mendelian syndromes, confirmed linkage to the region chromosome 9q22.32-31.1 using parametric (Heterogeneity Logarithm of Odds (HLOD) =1.23) and nonparametric linkage (NPL= 1.21) analyses. This linkage was found in 20% of the hereditary CRC families investigated. Subsequent analyses using more stringent thresholds, narrowed the region of susceptibility to just ~1.7 cM (NPL>2.4, p= 0.007).

A validation study was conducted in 2010 by Gray-McQuire and colleagues, which further confirmed the linkage of CRC to a susceptibility locus on chromosome 9q22. This region was fine-mapped using single nucleotide polymorphisms (SNPs) and moving window association analyses in 256 sibling colon cancer kindreds. A five-SNP haplotype was shown to be linked to the disease phenotype which was also in strong disequilibrium with another haplotype that lies in close proximity to *GALNT12*, which is located at chromosome 9q22.33. *GALNT12* encodes a glycosylation enzyme, UDP-N-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase 12 (GALNT12).

1.5.2 GALNT12: A Putative CRC gene

A study by Guda and colleagues (2009) identified a novel gene association with colon cancer. Somatic and germline variants were identified in the *GALNT12* gene in CRC cases. It was found that *GALNT12* deleterious variants inactivate normal functioning of the GALNT enzyme in initiating mucin type O-linked glycosylation. Guda and colleagues (2009) sequenced 30 microsatellite stable colon cancer cell lines and found 2 somatic variants, c.1023A>C, p.Glu341Asp and c.1436G>T, p.Cys479Phe, which were not found in the normal colon tissue. Through functional testing, it was found that the somatic mutations almost completely inactivated GALNT12 enzymatic activity (<3% of wild-type).

Guda and colleagues (2009) then analyzed the complete coding sequence of *GALNT12* in germline DNA samples from 272 colon cancer patients. Seven sequence variants were found among the colon cancer patients which were not found within the control group (192 individuals). Functional testing revealed that 6 of the 7 variants essentially knocked out enzymatic activity while the seventh variant reduced activity to 37% (Figure 1.3). The authors hypothesize that inactivity of this protein could increase the number of improperly glycosylated or unglycosylated proteins in colon cancer tumours. This study provides support for the notion that *GALNT12* pathogenic variants may contribute to the etiology of CRC.

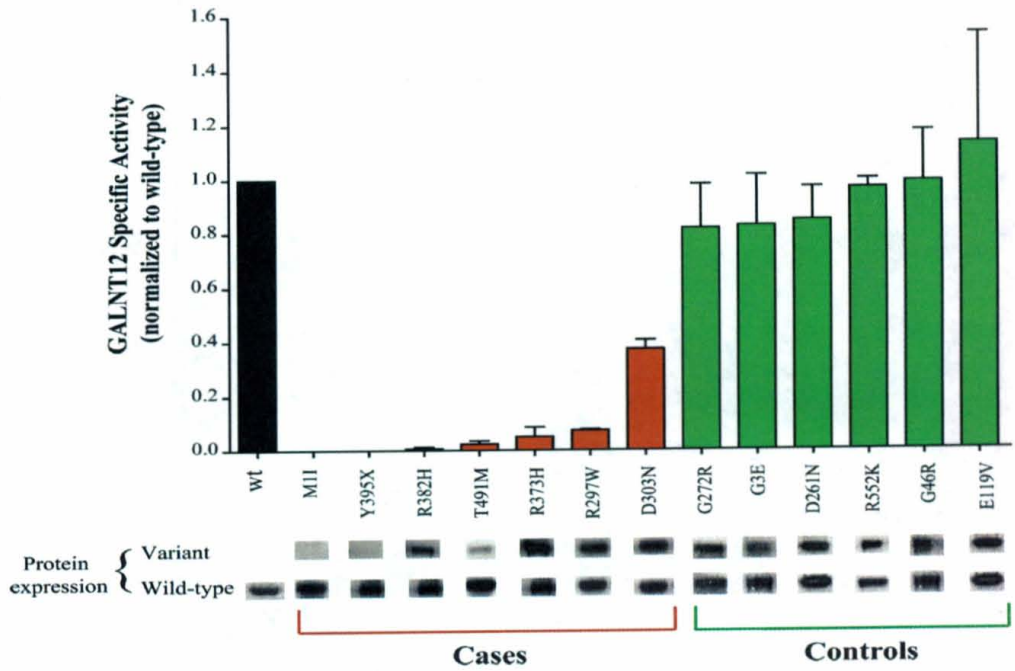

Figure 1.3. Characterization of *GALNT12* variants. *Upper*: specific enzymatic activity (normalized to wild-type) associated with variants found in germline DNA within CRC cases and controls. *Lower*: Protein expression associated with each variant. Adapted from Guda et al. (2009).

1.5.3 GALNT12 Function

GALNT12 has ten coding exons and is located on chromosome 9q22.33 (Figure 1.4). *GALNT12* encodes the enzyme *N*-acetylgalactosaminyltransferase type 12 (GALNT12) which belongs to a superfamily of *N*-acetylgalactosaminyltransferases (ppGalNac-Ts) (Hashimoto et al., 2011). The family of ppGaNTases are responsible for the transfer of a monosaccharide *N*-acetylgalactosamine (GalNAc) from UDP-GalNAC sugar donor to the hydroxyl group of a serine or threonine of the mucin protein (Hashimoto et al., 2011). This enzymatic monosaccharide transfer takes place in the Golgi and is considered the first step in mucin O-linked glycosylation (Ten Hagen et al., 2003). While the tertiary structure of the ppGalNac-Ts has not yet been determined, it has been suggested that all GALNT enzymes share a common β-pleated sheet with α-helices on either end (Hagen et al., 1999). There are approximately 24 unique human isoforms within the *GALNT* family which are tissue specific and *GALNT12* is known to be highly expressed in the human colon (Ten Hagen et al., 2003; Guo et al., 2002).

Mucin glycoproteins are ubiquitously distributed throughout epithelial cell surfaces and are important for cell adhesion, proliferation, differentiation, immunity, aging and cancer development (Hashimoto et al., 2011). The colonic epithelium, along with other epithelial surfaces in the body, is protected by a mucus layer that consists mainly of mucin glycoproteins (Byrd & Bresalier, 2004). Abberant glycosylation in mucin proteins is a hallmark feature of colonic neoplasia (Brockhausen, 1999). The proper glycosylation of mucin proteins plays an important role in the preferred backbone structure of the protein and influences the capability of the subsequent glycosylation events (Hashimoto et al., 2011). The altered structure of mucins through

A

B

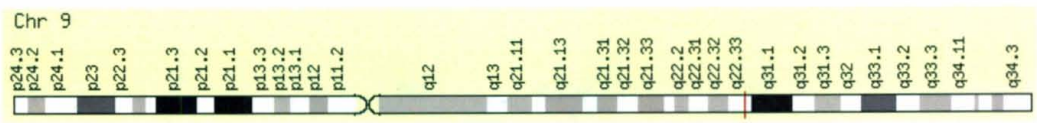

Figure 1.4. A: Coding nucleotide sequence *of GALNT12*. Nucleotide positions are numbered starting from the first nucleotide of the initiation codon. The predicted amino acid sequence is provided below. Arrows indicate the junction between introns and exons. The triangle identifies a truncated position. Adapted from Guo et al. (2002). **B**: *GALNT12* location on chromosome 9q22.33. Retrieved from http://www.genecards.org/cgi-bin/carddisp.pl?gene=GALNT12.

abberant glycosylation is associated with increased growth and survival of cells as well as an increased ability of cells to invade and metastasize (Brockhausen, 1999). Additionally, the proper structure of mucin proteins are important for interactions of potential ligands with cell surface receptors, which play a major role in immunity. Guda and colleagues (2009) found that there were increased levels of unglycosylated MUC1 proteins in 3 colon tumours compared to normal colon mucosa; although the expression of MUC1 proteins in the colon tumour was normal. Studies are ongoing to identify O-linked carbohydrate structures of mucins in CRC, which may be useful as diagnostic and prognostic tools as well as targets for cancer vaccines (Byrd & Bresalier, 2004).

Molecular cloning and characterization of *GALNT12* was carried out by Guo et al. (2002). They conducted a Quantitative real time PCR assay which revealed that *GALNT12* was mainly expressed in the tissues of digestive organs, including the stomach, small intestine and colon (See Figure 1.5).

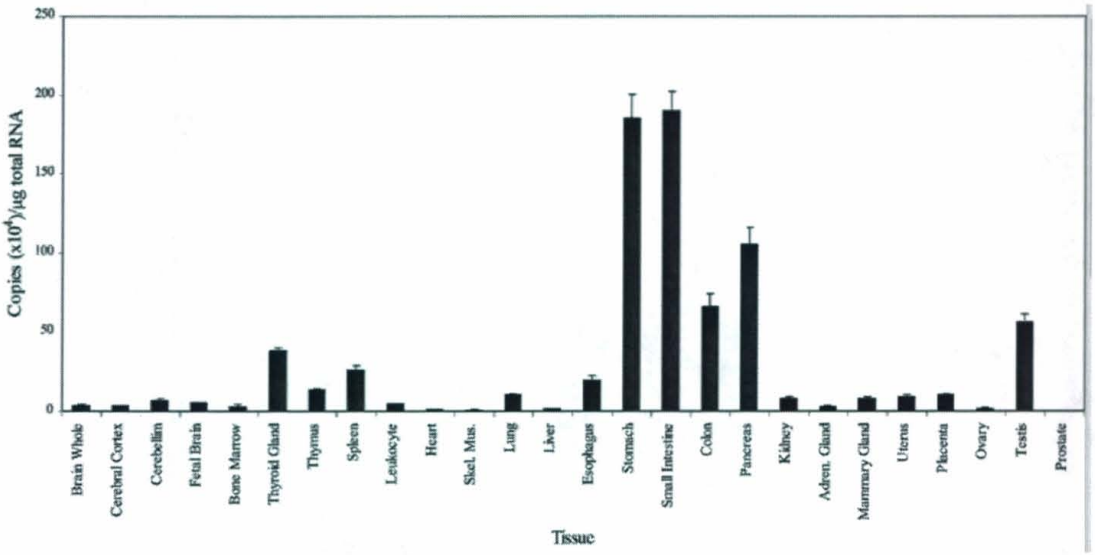

Figure 1.5. Expression of *GALNT12* in different tissues determined by Quantitative real time PCR. *GALNT12* is expressed the most in the digestive tissues including the stomach, small intestine and colon. Adapted from Guo et al. (2002).

Chapter 2

2.0 Investigating the Genetic Component of Colorectal Cancer in Newfoundland: A Clinical and Population-Based Study

2.1 Colorectal Cancer and the Newfoundland Population

The province of Newfoundland and Labrador (NL) has the highest incidence rates of CRC nationwide (Canadian Cancer Statistics, 2011), with 67 individuals per 100,000 affected compared to the national average of 49 individuals affected per 100,000 (Canadian Partnership Against Cancer, 2009). The mortality rate associated with CRC in Newfoundland is twice that in British Columbia, reflecting the high incidence rates in the NL population (Canadian Cancer Society, 2011).

The population structure of NL is unique in that it consists of numerous founder populations. This is due to the small number of early settlers to the province in the mid - 1700s. The first settlers were mostly Roman Catholics from Ireland or Protestants from England (Mannion, 1977). The island was initially populated with approximately 20,000-30,000 immigrants who settled along the coastline establishing out-port fishing communities. From 1760 to present, the population grew to over 500,000 due to natural expansion with over 50% of the population still living in rural communities of 2,500 people or less (Statistics Canada, 2010). The geographic isolation between communities and the religious pressures within each community produced several pockets of genetically isolated populations, some of which have been known to have founder mutations (Rahman et al., 2003). The founder effect occurs when a small group from a larger population forms a new colony. Any mutations that were present within the

individuals that branched off from the larger population will likely become established in the new population after several generations and these mutations will be present at a higher frequency in the smaller population compared to the original population. Founder mutations have been identified in a number of Newfoundland communities for several disorders including; Bardet-Biedl syndrome (Young et al., 1999), Hemophilia A (Xie et al., 2002) Multiple Endocrine Neoplasia type 1 (Olufemi et al., 1998) and CRC (Spirio et al., 1999). For example, a deleterious variant that affects the 3' splice acceptor site of *MSH2*, c.942+3A>T identified in Family C (an extensive hereditary CRC family), was also identified in 12 other families, which were all from the same geographical location (Green et al., 2002). Two large deletions of the *MSH2* gene were also identified; a deletion of exon 8 (c.1277-?_1386+?del, p.Lys427_Gln462>GlyfsX4) in 10 families and a deletion of exons 4 to 16 (c.646-?_2802+?del) in one very large family. Given its unique genetic structure, the Newfoundland population is an ideal resource for studying genetic diseases and identifying associated novel mutations.

In order to better understand the epidemiological, genetic and environmental factors associated with the high incidence of CRC in the Newfoundland population, the Newfoundland Colorectal Cancer Registry (NFCCR) was created. Individuals diagnosed with CRC from 1999 to 2003 were approached to be included in the registry (for more details see Materials and Methods-study population). This population-based cohort was assessed for familial risk of CRC based on FAP criteria, AC-I, AC-II, and Bethesda criteria outlined by Green et al. (2007) (see Table 2.1). It was found that 4.6% of this population met the high risk AC-II or FAP criteria; 43% met the intermediate risk

Table 2.1 CRC Risk Assessment

Family Risk	Category	Criteria	Comments
High–risk	FAP	Known FAP or AFAP	Mutation detected in *APC* gene or, clinical diagnosis of FAP or AFAP in proband's family
High-risk	H1	Family meets AC-I	Proband is not necessarily in the AC-I triad.
High-risk	H2	Family meets AC-II	Tumours recognized by the Amsterdam II criteria: colorectal, endometrial, small bowel, ureter, renal pelvis
Intermediate Risk	B1-B5	Proband meets revised Bethesda guidelines	Tumours: colorectal, endometrial, stomach, ovarian, pancreas, ureter and renal pelvis, small bowel, biliary tract, brain (glioblastoma), sebaceous cell adenomas and keratoacanthomas
Low-risk		Neither high or intermediate risk	

Adapted from Green et al. (2007).

Bethesda criteria and 30% of all the patients had a first- degree relative diagnosed with CRC (Green et al., 2007).When indicators of familial risk were compared to 13 other population-based studies worldwide, these indicators were significantly higher for the Newfoundland population (Green et al., 2007).

Woods et al. (2005) investigated tumours from high risk CRC families for deficiencies in MLH1, MSH2, and MSH6 proteins and DNA microsatellite instability. They found that 50% of the population-based families which fulfilled the high risk AC-I/II did not have MMR deficiencies. In addition, 76% of the high risk AC-I/II families referred to the Provincial Medical Genetics Program (PMGP) to did not have causal variants identified in the tested MMR genes.

Investigations carried out using the NFCCR have indicated that over 50% of the families meet the high or intermediate risk criteria (Green et al., 2007), suggesting that there is a high incidence of familial CRC in the Newfoundland population. Studies have also indicated that the high rate of familial CRC may be attributed to novel unidentified genetic factors (Woods et al., 2005; Woods et al., 2010). Considering the high incidence of unexplained familial CRC in the province of Newfoundland and several studies supporting the involvement of *GALTN12* in CRC, we investigated the involvement of *GALNT12* variants in the Newfoundland CRC population.

2.2 Objectives

The purpose of this study was to investigate the involvement of *GALNT12* causal variants in hereditary CRC families from Newfoundland. By better understanding the relationship between *GALNT12* causal variants and the onset of CRC, treatment and screening programs for this disease can become more specific and effective. There were two main objectives of this study.

1) To screen for variants in *GALNT12* in hereditary CRC families by performing direct DNA Sanger sequencing in 129 probands referred to the Provincial Medical Genetics Program.
2) To determine the causality of *GALNT12* variants found in probands by performing segregation studies in family members and also screening population-based CRC cases and controls.

The impact of these variants was assessed by sequencing three different groups; (1) affected and unaffected family members of the probands from the clinical cohort; (2) a replication group from the NFCCR to determine the CRC population-based frequency of the variants, and (3) a healthy control group to determine the population based frequency of the identified variants. Bioinformatic tools and functional test results by Guda et al. (2009) were also used to help determine the impact of variants on protein function.

Chapter 3

3.0 Materials and Methods

3.1 Study Population

CRC patients included in the study came from two separate cohorts: clinic-based and population-based registries. The first cohort included 129 individuals diagnosed with CRC, who were referred to the Provincial Medical Genetics Program by their physician. Venous blood was drawn from the proband and family members of each family and whole genomic DNA was extracted. The second cohort included 481 individuals who were ascertained through the Newfoundland Colorectal Cancer Registry (NFCCR) that had not previously been referred to the Provincial Medical Genetics Program. Subjects in the NFCCR were recruited during the period of 1999-2003 and were diagnosed with CRC before the age of 75 in Newfoundland and Labrador. DNA samples, paraffin-embedded tumour samples, family history and access to medical records were obtained from each consenting individual. Control DNA samples were obtained from volunteers by random digit dialing. These subjects had no personal or family history of epithelial cancer, they provided a blood sample, provided family history and consented to be included in research studies. This project was approved by the Research Ethics Board of Memorial University of Newfoundland.

3.2 Polymerase Chain Reaction

Polymerase chain reaction (PCR) was conducted to amplify all coding sequence and flanking intronic sequences of *GALNT12*. Forward and reverse primer sequences were obtained from Guda et al. (2009). Primers were manufactured by Eurofins MWG

Operon (Huntsville, Alabama). Stock primers were diluted to a final concentration of 100 ng/µl (see appendix C for primer sequences). Positive controls were used to obtain optimal conditions for the PCR reaction using control DNA at 100 ng/µl.

Approximately 1.0 µl of genomic DNA from a stock concentration of 100 ng/µl was added to a 96-well 0.2 ml PCR reaction plate. For reactions amplifying exons 2-10 a cocktail containing 10.75 µl of deionized water, 1.5 µl of 10X PCR reaction Buffer ($MgCl_2$), 0.3 µl of 100 ng/µl dNTPs (deoxyribonucleotide triphosphate), 0.5 µl of 10 µM forward primer, 0.5 µl of 10 µM reverse primer, 0.75 µl of 50 mM MgCl and 0.2 µl of Platinum Taq Polymerase was prepared and added to the 96-well 0.2 ml PCR reaction plate (materials for the cocktail were obtained from Life Technologies, Burlington, ON). Negative controls were tested using 1.0 µl of deionized water instead of DNA. The plate was then centrifuged and incubated in a Biometra Thermocycler (Biometra GmbH, Goettingen, Germany) or Eppendorf Thermocycler (Eppndorf, Hamburg, Germany). The thermocycler conditions were as follows: an initial cycle of denaturation for 5 min at 95.0 °C. This was followed by 30 s of denaturation at 95 °C, annealing of primers for 30 s at 60.0 °C and extension of DNA strands for 30 s at 72.0 °C for 25 cycles. This was followed by extension of strands for 7 min at 72.0 °C and then a hold at 4 °C.

Due to difficulty in amplifying exons 1 and 10d, a different protocol was used. A cocktail containing 22 U/ml *Thermococcus* species KOD thermostable polymerase, anti-KOD antibodies, 66 mM Tris-SO_4 (pH 8.4), 30.8 mM $(NH_4)_2SO_4$, 11 mM KCl, 1.1 mM $MgSO_4$, 330 µM dNTPs, AccuPrime™ proteins, stabilizers 1.0 µl of 10 µM forward primer, 1.0 µl of 10 µM reverse primer.

For exon 1 thermocycler conditions were as follows: an initial cycle of denaturation for 3 min at 95.0 °C, followed by 30 s of denaturation at 95 °C, annealing of primers for 30 s at 55.0 °C and extension of strands for 1 min at 72.0 °C for 35 cycles. This was followed by extension for 10 min at 72.0 °C and then a hold at 4 °C. For exon 10d the thermocycler conditions were as follows: an initial cycle of denaturation for 2 min at 95.0°C. This was followed by 20 s of denaturation at 94 °C, annealing of primers for 45 s at 60.0 °C and extension for 45 s at 72.0 °C for 14 cycles. This was followed by 20 s of denaturation at 94 °C, annealing of primers for 45 s at 50.0 °C and extension for 30 s at 72.0°C for 32 cycles completed with a hold at 4 °C.

To confirm that the PCR was producing the appropriate amplicon, the products were added to a 2% agarose gel which contained 1g of Ultrapure Agarose (Life Technologies, Burlington, ON), 50 ml of Tris/Borate/EDTA (TBE) buffer and 3.8 µl of SYBR®Safe (Life Technologies, Burlington, ON). Approximately 3.0 µl of PCR product was mixed with 3.0 µl of dye and 1.0 µl of DNA ladder was mixed with 3.0 µl of dye. These mixtures were added to the agarose gel. The gel was then subjected to electrophoresis (120 V for 25 min). An image of the gel was then captured using a Kodak UV imager (Kodak, Rochester, New York).

3.3 Polymerase Chain Reaction Purification

PCR products were purified using Exo Nuclease I (ExoI) and Shrimp Alkaline Phosphatase (SAP). ExoI was used to digest any excess primer left in the PCR product. SAP was used to dephosphorylate free deoxynucleotide triphoshates (dNTP's) from the product. A cocktail containing 7.5 µl of deionized water, 0.5 µl of ExoI (10 U/ µl) and

0.5 µl of SAP (1 U/ µl) was added to a 96 well 0.2 ml PCR plate. Approximately 8.0 µl of PCR product was added to the mixture and incubated at 37°C for 30 min followed by a 15 min incubation at 80°C to inactivate the enzymes.

3.4 Sequencing Reaction

Bi-directional sequencing was performed for each of the coding regions and the exon/intron boundaries of the *GALNT12* gene for the clinical cohort. Bidirectional sequencing was carried out in only exons 4 and 6 for the population-based cohort as variants identified in the clinic-based cohort occurred in these exons. A cocktail containing 2.0 µl of BigDye Terminator v3.1 Cycle Sequencing 5X Buffer, 0.5 µl of BigDye Terminator v3.1 Cycle Sequencing Mix (Applied Biosystems, Foster City, California), 0.67 µl of primer (1.6 pmol/µl), and 14.83 µl of deionized water was prepared. Approximately 18.0 µl of the cocktail and 4.0 µl of purified PCR product was added to a 96 well 0.2 ml PCR plate. The plate was then incubated at 96°C for 1 min followed by 25 cycles of 96 °C for 10 s, 50°C for 5 s and 60 °C for 4 min.

3.5 Sequencing Reaction Purification

The sequencing reaction was then stopped by adding 5 µl 1.25mM ethylenediaminetetraacetic acid (EDTA) followed by 65 µl of 95% ethanol solution. The plate was vortexed and placed in the dark overnight to allow the sequencing product to precipitate out of solution. The plate was then centrifuged for 45 min at 3000g. The plate was then inverted and centrifuged for 30 s at 350 rpm to allow the ethanol to decant. The samples were washed with 150 µl of 70% ethanol solution and the plate was centrifuged at 3000g for 15 min. The plate was inverted again and placed in the centrifuge at 350 rpm

for 30 s to remove any excess supernatant. Samples were resuspended in 18 µl of HiDi formamide by vortexing. The plate was then centrifuged and the sequencing product was denatured by incubation at 95 °C for 2 min. The caps were then replaced with a Septa seal and placed in the ABI Prism 3130 or 3730 Automated Genetic Analyzer for automated sequencing (Applied Biosystems, Foster City, California).

3.6 Sequencing Analysis

Data obtained from the automated genetic analyzer was examined for quality control using Sequencing Analysis v5.2 software and also by manual analysis. The target quality versus fair and poor qualities are shown in Figure 3.1. Sequence of good quality shows high sequenced peaks with minimal background peaks. As the quality moves to fair and poor there is an increase of background noise that is less distinguishable from the main sequence peaks, increasing the difficulty of making accurate nucleotide calls. Fair and poor quality sample sequences were only used in conjunction with a good quality sequence for that sample in the opposite direction.

The sequence data from the genetic analyzer was investigated for variants by comparing it to a reference sequence obtained from the Ensembl database (http://ensembl.org/index.html). Sequencher v4.9 software was used to align the sample sequence with the reference sequence to identify variants. Only variations that occurred in both the forward and reverse sequences were considered to be genuine. Potential pathogenic variants were confirmed by repeating the reaction using a separate aliquot of DNA from the same individual for quality control purposes

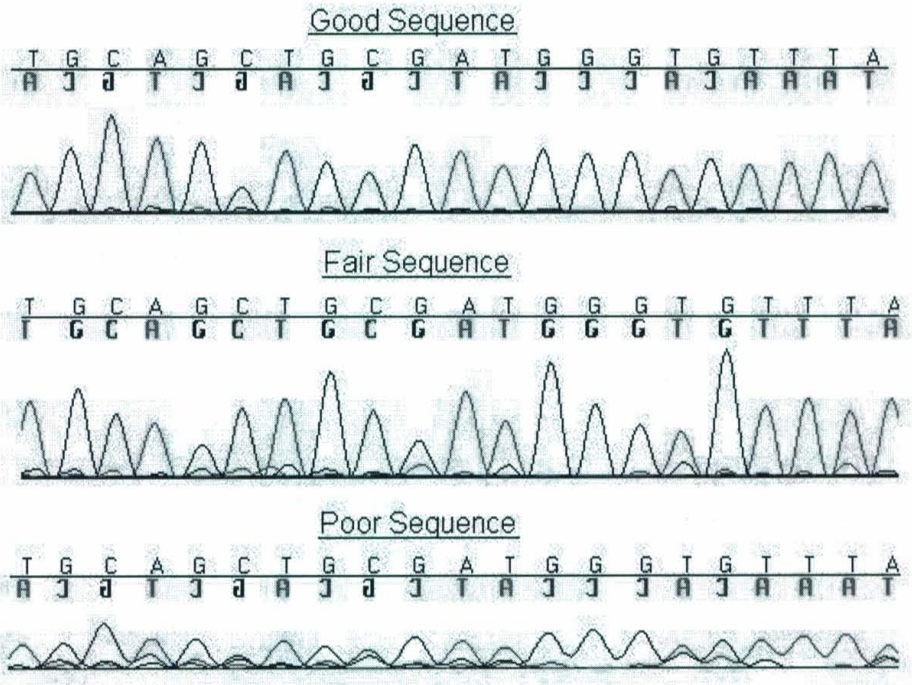

Figure 3.1 Manual examination of data quality. Adapted from Kamel (2010).

3.7 Segregation Analysis

SImplified rapid Segregation Analysis (SISA) was used to assess the probability of co-segregation by chance of *GALNT12* variants in families that showed segregation. (Møller et al., 2011). SISA assesses penetrance and expression of genetic variants that have unknown pathogenic significance based on a number of assumptions; (1) the variant is introduced only once into that family; (2) the probability of recombination is zero; and (3) a family member that occurs between two members that are known to be carriers of the variant is considered to be an obligate carrier. The *probability of cosegregation by chance* is calculated by $(1/2)^n$, where n equals the number of informative meiosis (number of people included in the carrier set minus one). *Expression* (fraction of individuals that have a specific phenotype among all affected individuals) is simply calculated as the number of persons with one (or some) of the phenotypic characteristics considered to be associated with affected status, divided by the total number of individuals who are affected. *Penetrance* is calculated by the number of affected carriers divided by the number of carriers. Calculating a low probability for co-segregation by chance and high penetrance may indicate that the variant of unknown significance is causing the phenotype. Calculating a high probability of cosegregation by chance and a low penetrance may indicate that the variant is not associated with the observed phenotype. For a more detailed explanation, see Møller et al. (2011).

3.8 Determining Pathogenicity of Identified Variants

The population frequency of all variants, identified though sequencing, within the exons and exon/intron boundaries, was determined using the Ensembl database. Ensembl

database is linked to the global SNP database, dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), which reports human population frequencies of these variants. Population frequencies were noted (when available) for all variants that were previously identified in dbSNP based on the 1000 Genomes Project (http://www.1000genomes.org/).

All variants with a population frequency of less than 1% were investigated within the family for segregation patterns with the CRC phenotype. All non-synonymous variants affecting amino acid sequence were investigated for pathogenicity using bioinformatic software including PolyPhen (Adzhubei et al., 2010; http://genetics.bwh.harvard.edu/pph2/), SIFT (Ng & Henikoff, 2001; http://sift.jcvi.org/) and Panther (Thomas & Kejariwal, 2004; http://www.pantherdb.org/). Align GVGD was also used to predict pathogenicity of variants identified in this study. Align GVGD is an expansion of the Grantham chemical difference matrix (Grantham, 1974) which takes into account the biophysical characteristics of amino acids and also uses protein multiple sequences alignments to estimate if protein variants are deleterious or neutral (Tavtigan et al., 2005; Mathe et al., 2006). For variants previously reported by Guda and colleagues (2009), enzymatic function was based on previously reported functional testing. The Newfoundland population based frequency of putatively pathogenic variants was investigated by sequencing exons of age-matched controls.

3.9 Statistical Analysis

Statistical analysis was carried out to analyze differences between the frequencies of potential pathogenic variants in the case and control group. Fisher's exact test was

employed using QuickCals online software (http://www.graphpad.com/quickcalcs/index.cfm).

3.10 Criteria for Affected Family Members

While all the probands included in both cohorts have been diagnosed with at least a single case of CRC, relatives of the proband included in the disclosed family history were considered affected if they had been diagnosed with CRC or at least one adenomatous polyp. Adenomatous polyps are considered a precursor to CRC as these polyps often progress into a malignant form. It is also important to note that, due to increased screening in individuals belonging to hereditary CRC families, affected family members are more likely to have adenomatous polyps removed before they develop into cancer. Therefore, it is important to include individuals with adenomatous polyps in the affected phenotype.

Chapter 4

4.0 Results

4.1 Clinically Referred Families

There were 129 probands from families referred to the PMGP included in this study. Approximately 68% of the cohort met the Bethesda criteria, 21% met the Amsterdam I/II criteria and 11% did not meet either criteria and were considered low risk. Of the 129 probands, DNA from 120 cases was successfully sequenced in all coding regions and intron/exon boundaries of *GALNT12*. Four individuals were found to harbour putatively deleterious variants. A number of putatively non-pathogenic variants were also identified. All variants found within this cohort are reported in Table 4.1.

4.1.1 Putatively Pathogenic Variants

Two potential pathogenic variants were identified in four different probands. Three cases were found to harbour the c.907G>A (p.Asp303Asn) variant, which was previously reported by Guda and colleagues (2009). A novel variant, c.1187A>G (p.Tyr396Cys) was identified in one individual. A summary of the putatively pathogenic variants identified in *GALNT12* is shown in Table 4.2.

The variant, c.907G>A (p.Asp303Asn), was investigated using bioinformatic tools such as PolyPhen, SIFT and Panther. PolyPhen predicts this variant to be "probably damaging" with a score of 0.999 (a score of 0.0 being neutral and a score of 1.0

Table 4.1 Germline variants in *GALNT12* identified in clinically based probands.

Variant	Exon	Frequency		
		Cases	Controls	Population[1]
c.136G>A (p.Gly46Arg) (rs10987768)	1	0.008	ND	ND
c.356A>T (p.Glu119Val) (rs10987769)	1	0.083	ND	0.018
c.781G>A (p.Asp261Asn) (rs41306504)	4	0.016	0.012	0.008
c.907G>A (p.Asp303Asn)	4	0.025	0.00	ND
c.1187A>G (p.Tyr396Cys)	6	0.008	0.00	ND
c.1497C>T (p. =) (rs35632007)	9	0.008	ND	0.026
c.1707G>C (p. =) (rs2273846)	10	0.046	ND	0.157
c.1734G>A (p. =)	10	0.016	ND	ND
c.*171A>G (rs2773847)	10	0.033	ND	0.157
c.*421G>A (rs2273848)	10	0.083	ND	0.064

[1]Data sourced from dbSNP [http://www.ncbi.nlm.nih.gov/projects/SNP/] via Ensembl [http://www.ensembl.org]; ND: not determined

Table 4.2 Putatively pathogenic *GALNT12* variants in clinically based probands with clinical information.

Study #	Variant	Proband Age of Diagnosis	Location of Cancer	MSI[1] Testing	IHC[2] Testing	Family[3] Risk
1117	c. 907G>A, p.Asp303Asn	72	Rectum	MSS	intact	B24
20444	c.907G>A p.Asp303Asn	35	Rectum	ND	ND	B45
20896	c.907G>A p.Asp303Asn	71	Transverse Colon	MSS	intact	B25
1728	c.1187A>G p.Tyr396Cys	21	Cecum	MSS	intact	B1

[1]MSS: microsatellite stable; ND: not determined. [2]IHC testing was conducted using the MLH1, MSH2 and MSH6 proteins. [3]Bethesda criteria: 1, CRC diagnosis less than 50 years of age; 2, synchronous or metachronous tumours (colorectal or HNPCC); 4, one 1st degree relative with HNPCC related tumour under 50 years of age; 5, two 1st or 2nd degree relatives with HNPCC related tumours at any age. Family risk criteria are outline in Green et al. (2007).

representing deleterious changes). SIFT predicts this change to be "not tolerated" and Panther reported a subPSEC score of -4.33 and a $P_{deleterious}$ score of 0.791 for the p.Asp303Asn change. The subPSEC score in Panther represents the "negative logarithm of the probability ratio of the wild type (WT) and mutant amino acids at a particular position" (Brunham et al., 2005). These scores are continuous values that range from 0 (mostly likely benign) to -10 (most likely pathogenic). The $P_{deleterious}$ is the probability of functional impairment of a specific amino acid change (Brunham et al., 2005). See Table 4.3 for a summary of the bioinformatic data obtained. The Grantham chemical difference matrix was also employed to estimate the pathogenicity of this variant, c.907G>A (see Materials and Methods section 3.8). The Align GVGD classification for this amino acid change, p.Asp303Asn, is C15 (Table 4.3). Scores for amino acid changes range from C0, less likely to interfere with the function of the protein, to C65, most likely to interfere with protein function.

The variant, c.1187A>G (p.Tyr396Cys), was not identified by Guda et al. (2009) and therefore functional testing has not been previously conducted. We conducted bioinformatic analysis on this variant and it was predicted to be pathogenic (PolyPhen, SIFT and Panther). PolyPhen predicted this variant as "probably damaging" with a score of 0.999. SIFT predicts this change as "tolerated" and Panther calculates a subPSEC score of -9.28 and a $P_{deleterious}$ of 0.998 (Table 4.3). The Align GVGD classification for this amino acid change, p.Tyr396Cys, is C65 (Table 4.3).

Table 4.3 Bioinformatic analysis of nonsynonymous *GALNT12* variants identified in clinical probands.

		Bioinformatic Tools			
Variant	Exon	SIFT[1]	PolyPhen[2] (Prediction, $P_{deleterious}$)	Panther[3] (subPSEC, $P_{deleterious}$)	Align GVGD[4]
c.136G>A (p.Gly46Arg)	1	TD	Unknown	-2.99, 0.498	C65
c.356A>T (p.Glu119Val)	1	TD	PsD, 0.788	-4.62, 0.865	C65
c.781G>A (p.Asp261Asn)	4	TD	PrD, 0.997	-3.35, 0.586	C15
c.907G>A (p.Asp303Asn)	4	NT	PrD, 0.999	-9.28, 0.998	C15
c.1190A>G (p.Tyr396Cys)	6	NT	PrD, 0.999	-4.33, 0.791	C65

[1]TD: tolerated; NT: not tolerated. [2]PsD: possibly damaging; PrD: probably damaging. [3] subPSEC: the "negative logarithm of the probability ratio of the wild type and mutant amino acids at a particular position" (Brunham et al., 2005). These scores are continuous values that range from 0 (mostly likely benign) to -10 (most likely pathogenic). [4]Classifications are listed from most likely to interfere with protein function to least likely; C65 (most likely), C55, C45, C35, C25, C15, C0 (less likely).

The variant, c.907G>A (p.Asp303Asn), was identified in a Newfoundland family of European descent (family 20444). This family fulfills the Revised Bethesda criteria: however, it seems likely that this family may have met the high-risk Amsterdam I criteria if the youngest generation had not been receiving high-risk colonoscopy screening (Figure 4.1). The point mutation, c.907G>A (p.Asp303Asn), causes an amino acid change from aspartic acid to asparagine (p.Asp303Asn). This variant has been previously reported by Guda and colleagues (2009) to reduce enzymatic activity to 37% of WT function. The female proband (PID 6) was diagnosed with cancer of the rectum at age 35 and skin cancer at age 82. All four children of the proband were negative for the c.907G>A variant. The proband's sister (PID 4) also carried the point mutation and was diagnosed with cervical cancer at age 37 and colon cancer at age 61. Six of her eight children also have the variant; no DNA was available for a ninth child. Four of the six carriers have presented with colonic polyps (PID 7, 8, 10, 12) while two of the carriers did not have polyps. Two other siblings were non-carriers and have not presented with polyps or CRC (PID 14, 15). Of the four carriers that have been diagnosed with polyps, two had adenomas (PID 7, 12), one had hyperplastic polyps (PID 8) and the other individual's polyp type is unknown (PID 10). One individual who is negative for the variant had a small cecal polyp which was not recovered for pathological analysis (PID 13; Table 4.4).

Unfortunately, adequate size tumour tissue was not available from the proband or her sister for microsatellite instability and immunohistochemistry (IHC) analysis. The *MLH1, MSH2* and *MSH6* genes were sequenced in the proband and her sister. No

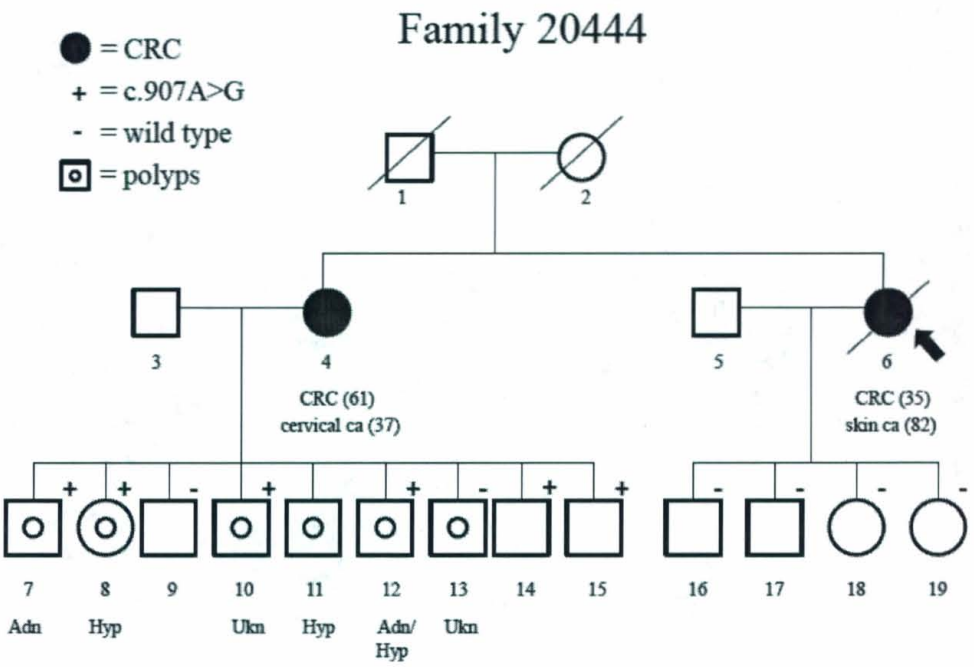

Figure 4.1 Partial pedigree for family 20444 with the segregating c.907G>A, p.Asp303Asp *GALNT12* variant. Arrow indicates proband. Hyp: hyperplastic polyp; Adn: adenomatous polyp; Ukn: polyp type unknown.

Table 4.4 Medical history of clinical family members with probands harbouring a putatively pathogenic variant in *GALNT12*.

Study #	PID[1]	Cancer Diagnosis[2]	Polyp Data[3]
20444	4	Cervical Cancer 37yrs CRC (descending) 61 yrs.	NA
20444	6	CRC (rectum) 35 yrs. Skin cancer 82 yrs.	NA
20444	7	NA	2 TAs at 43 yrs. Mult. TAs at 44/45 yrs. 1 TA at 54 yrs. 2 TAs at 60 yrs.
20444	8	NA	1 HYP at 60 yrs.
20444	10	NA	Small polyps (age unknown, type unknown).
20444	11	NA	3 HYPs at 37 yrs.
20444	12	NA	7 TAs at 51 yrs. 1 TA and 1 HYP at 53 yrs.
20444	13	NA	1 small cecal polyp at 47 yrs. Normal colonoscopy at 49 yrs.
20896	3	3 CRCs; 1 at 72 yrs. and 2 at 79 yrs.	Mult. TAs and HYPS between 73 and 81 yrs.
20896	7	2 CRCs; 1 at 77 yrs. and 1 at 83 yrs.	1 pedunculated polyp and 2 sessile polyps (age unknown)
20896	8	Upper GI "Ca of stomach" in "late 70s" Mass in transverse colon at 82 yrs.	NA

20896	11	NA	1 TA with severe dysplasia at 53 yrs. Several small polyps and 2 HYPs (age unknown).
20896	12	Duke's A rectal carcinoma at 57 yrs.	2 TAs at 35 yrs. 1 TA at 60 yrs. (moderate dysplasia) 3 TAs at 61 yrs. 3 TAs at 62 yrs.
20896	13	NA	Normal colonoscopy at 47 yrs.
20896	14	NA	2 HYPs at 54 yrs. 2 large TAs on stalks with moderate dysplasia at 45 yrs.
20896	18	Duke's C carcinoma at 57 yrs. (in preexisting TA)	TA at 57 yrs.
1117	12	CRC at 47 yrs.	NA
1117	18	2 CRCs at 72 yrs. Ca skin at 74 yrs.	NA
1117	19	CRC at 67 yrs.	NA
1117	23	CRC at 62 yrs.	NA
1117	5	Ca stomach at 68 yrs.	NA
1117	6	CRC *in situ* at 55 yrs.	NA
1117	7	CRC at 70 yrs.	NA
1117	8	CRC at 70 yrs.	NA
1117	15	CRC at 21 yrs. (Ca appendix?)	NA

[1] PID: personal identification number; [2]NA: not applicable, CRC: colorectal cancer, TA: tubular adenoma, HYP: hyperplastic polyp

deleterious variants were identified in these MMR genes. The *MUTYH* gene was not investigated in this proband.

A second Newfoundland family was identified with the same c.907G>A (p.Asp303Asn) variant. The male proband (PID 3) of family 20896 presented with multiple adenomatous and hyperplastic polyps, was diagnosed with a colon tumour at age 72, and two synchronous colon tumours at age 79 (Figure 4.2; Table 4.5). The proband had six children, three of whom took part in the study (PID 12,13,14). One daughter (PID 12) who was found to have the variant, was diagnosed with rectal cancer at age 57, and presented with three adenomas at 61 years and an additional three adenomas at 63 years. A second daughter (PID 14) was also found to carry the c.907G>A variant and she presented with two colonic adenomas with moderate dysplasia at age 45. A third daughter (PID 13) was found to be WT for the variant and presented with two hyperplastic polyps at 54 years. The proband had twin sisters who were both diagnosed with cancer. One twin sister (PID 7) was diagnosed with cancer of the cecum at age 77 years and a second colon cancer at age 83. This woman had a son (PID 18) who was diagnosed with colon cancer at age 57, and was found to be WT for the c.907G>A (p.Asp303Asn) variant. The other twin sister (PID 8) had cancer of the stomach (late 70s) and also presented with a mass in her transverse colon at 82 years. This mass was not surgically removed nor was it biopsied. This family fulfills the Revised Bethesda Criteria and may have met the high-risk Amsterdam criteria if they were not undergoing "high-risk" screening for the disease. The proband of this family does not harbour any disease causing variants in the *MUTYH* gene and the proband's tumour sample is microsatellite stable (MSS) and intact for IHC of MMR proteins (Woods et al., 2010). The pedigree is presented in Figure 4.2.

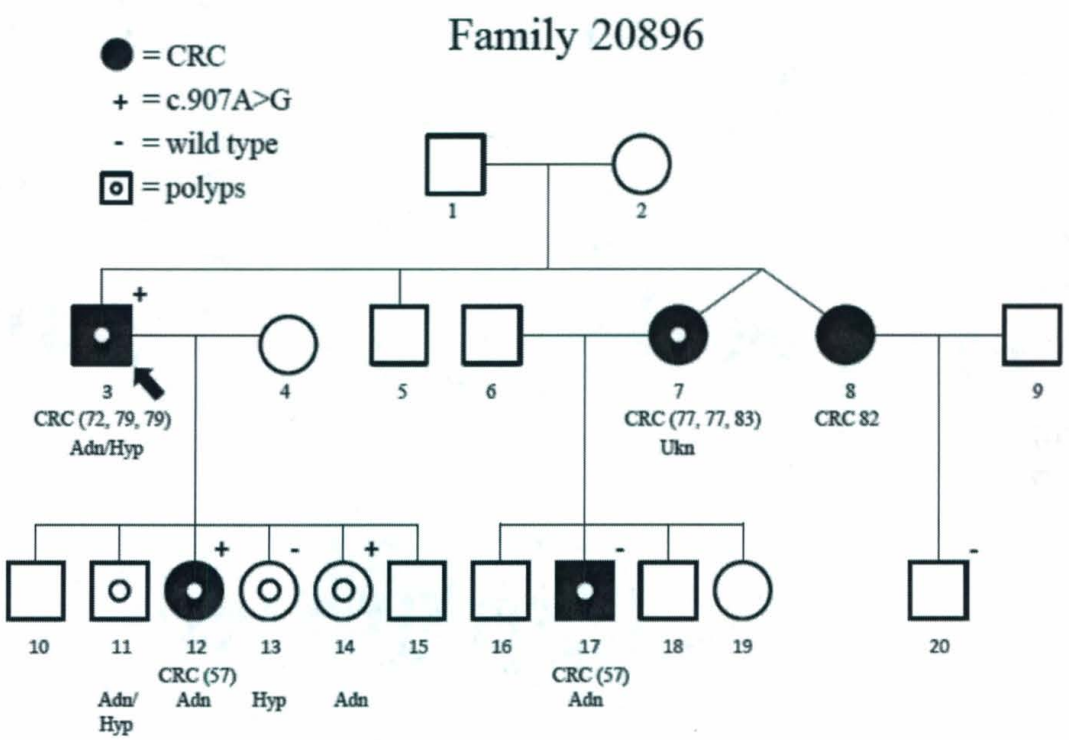

Figure 4.2 Partial pedigree for family 20896 with the partially segregating c.907G>A, p. Asp303Asn *GALNT12* variant. Arrow indicates proband. Hyp: hyperplastic polyp; Adn: adenomatous polyp; Ukn: polyp type unknown.

The medical histories for family 20444 and 29896 are presented in Table 4.4.

SISA was conducted on families 20444 and 20896 to determine the probability of co-segregation by chance of the c.907G>A variant with CRC and/or adenomatous polyps. The probability of co-segregation of c.907G>A with the disease phenotype by chance, for both pedigrees combined, was 1.56%. Family 20444 had a 6.25% probability of co-segregation by chance and family 20896 had a 25% probability of co-segregation by chance.

A third proband harboring the c.907G>A (p.Asp303Asn) variant is from family 1117 and fulfills the Revised Bethesda Criteria (PID 18; Figure 4.3). The female proband had two synchronous CRCs at the age of 72 years and a subsequent skin cancer at the age of 74 years. The brother of the proband (PID 12) had CRC at age 47 years but no DNA was available for testing. Two male first cousins once removed of the proband (PID 19, 23) also were diagnosed with CRC. Bi-directional sequencing revealed that they did not carry the c.907G>A variant. The father of these two male first cousins once removed (PID 11) had no personal or family history of CRC. The proband of this family does not harbour any disease-causing variants in the *MUTYH* gene and the proband's tumour sample is MSS and intact for IHC of MMR proteins (Woods et al., 2010).

A novel variant, c.1187A>G (p.Tyr396Cys), was identified in an individual from family 1728 who was diagnosed with CRC at age 21years (Figure 4.4). While neither parent of the proband (PID 13, 14) was diagnosed with CRC, his paternal grandmother

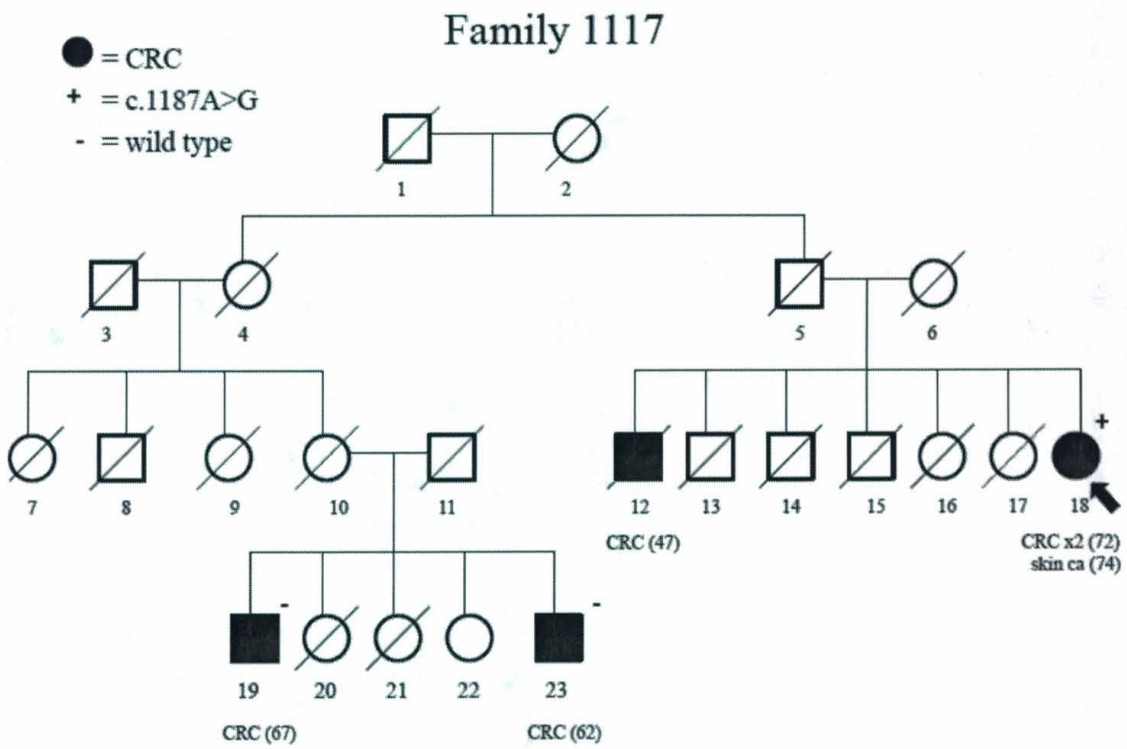

Figure 4.3 Partial pedigree for family 1117. No Segregation of c.907G>A, p.Asp303Asn *GALNT12* variant with CRC phenotype. Arrow indicates proband.

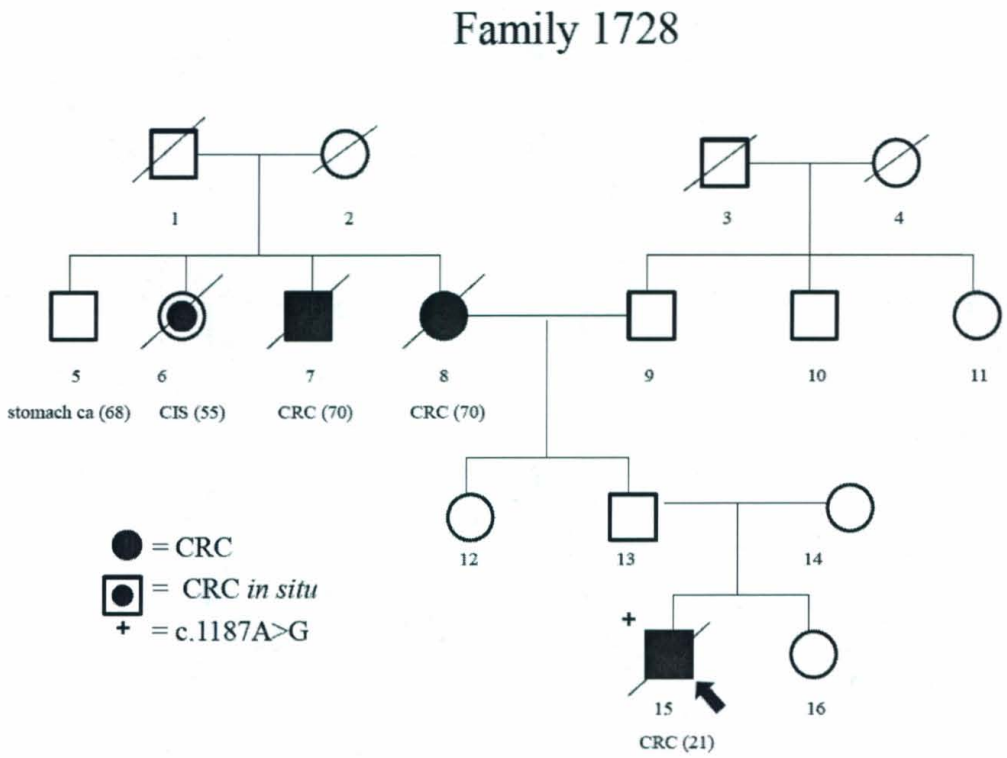

Figure 4.4 Partial pedigree for family 1728. The proband harboured the c.1187G>A, p.Tyr396Cys *GALNT12* variant. No other family members were available for testing. Arrow indicates proband; CIS refers to CRC *in situ*.

(PID 8) and her brother (PID 7) both had CRC at age 70 years. A paternal great aunt (PID 6) presented with a malignant polyp at 55 years and a second paternal great uncle (PID 5) was diagnosed with cancer of the stomach at age 68. Unfortunately, DNA from these individuals was not available for testing. The *MUTYH* gene was not investigated for pathogenic variants for this individual. The proband's tumour did not exhibit molecular characteristics consistent with MMR mutations (Woods et al., 2010).

4.1.2 Investigating Pathogenic Variants in Control Samples

Using 298 chromosomes from healthy individuals with no family history of cancer, we sequenced exons 4 and 6 to obtain the population frequency of both potentially pathogenic variants, c.907G>A and c.1187A>G. Neither variant was identified in any of the 149 individuals tested, p= 0.039 (Fisher's exact test).

4.1.3 Putatively Non-pathogenic Variants

Other *GALNT12* variants were also identified by sequencing the probands (Table 4.1). These variants were categorized as putatively non-pathogenic based on three conditions: 1) the population frequency of the variant was greater than 1%; 2) the variant was previously reported in the control group in Guda et al.(2009); and 3) the variant codes for a synonymous amino acid change.

The exon 1 variant, c.136G>A (p.Gly46Arg), was identified in 1 individual of the 120 probands. This variant has been previously reported in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=10987768); however, the population based frequency of this variant remains unknown. Guda and colleagues (2009) reported

that this variant reduces enzymatic activity to 99% of the WT function. This variant was also found only in the controls included in Guda et al. (2009). Panther predicts a subPSEC score of -2.99 and a $P_{deleterious}$ of 0.498 for this variant. PolyPhen could not predict a possible outcome for this amino acid change. SIFT predicts this amino acid change, p.Gly46Arg, as "tolerated". The Align GVGD classification for this change, p.Gly46Arg, is C65 (Table 4.3).

An exon 1 variant, c.356A>T, p.Glu119Val, was identified in 16/120 individuals (13.3%). This variant has been reported in 1.8% of the population based on the 1000 genomes project as reported in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=10987769). Guda and colleagues (2009) also found this variant in their control group and functional testing revealed the associated enzymatic activity to be 113% of WT activity. PolyPhen predicts this variant as "possibly damaging" with a score of 0.788. SIFT predicts the amino acid change p.Glu119Val as "tolerated". Panther predicts a subPSEC score of -4.62 for the p.Glu119Val change and a $P_{deleterious}$ of 0.865. The Align GVGD classification for this amino acid change, p.Glu119Val, is C65 (Table 4.3).

Two individuals (1.6%) harboured a variant in exon 4, c.781G>A, p.Asp261Asn. This variant has been reported in 0.8% of the population based on the 1000 genomes project as reported by dbSNP (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=41306504). This variant was also reported by Guda and colleagues (2009) in their control group with an enzymatic activity of 84.7% normalized to WT function. Three of the 298 (1.0%) control sample chromosomes sequenced in this study had the c.781G>A variant. PolyPhen predicts

p.Asp261Asn as "probably damaging" with a score of 0.997. SIFT predicts the change as "tolerated". Panther calculated a subPSEC score of -3.35 and a $P_{deleterious}$ of 0.586. The Align GVGD classification for this amino acid change, p.Asp261Asn, is C15 (Table 4.3).

A synonymous variant was identified in exon 9, c.1497C>T (p. =). One individual harboured this variant and it was previously reported in 0.5% of the population based on the 1000 genomes project as reported by dbSNP (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=35632007). This variant was not investigated using bioinformatic tools as it does not alter protein sequence.

Two synonymous variants were identified in exon 10, c.1707G>C and c.1734 G>A. The c.1707G>C variant has been previously reported to be in 15.7% of the population based on the 1000 genomes project as reported by dbSNP (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=2273846). This variant was identified in 10/120 (8.3%) of the cases sequenced. The c.1734G>A variant has not been previously reported, but is considered to be non-pathogenic based on the predicted synonymous amino acid. Two individuals harboured this variant. These variants do not alter the protein sequence and therefore were not investigated using bioinformatic tools.

Two variants were also identified in the untranslated region (UTR) of exon 10, c.*171A>G and c.*421G>A. The c.*171A>G variant has been previously reported as having a population based frequency of 15.7% based on the 1000 genomes project as reported by dbSNP (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=2273847). This variant was found in 6/120 cases (5.0%), with one individual homozygous for the change. The c.*421G>A is also a known variant and is reported to have a population based

frequency of 6.4% (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=2273848). This variant was found in 12/120 cases (10.0%) sequenced.

4.2 Newfoundland Colorectal Cancer Registry- Population Based Cases

After identifying two putatively pathogenic variants in exon 4 and exon 6 of *GALNT12* by investigating Newfoundland hereditary CRC families, cases from the NFCCR were sequenced for exons 4 and 6 to identify the frequency of these variants in the Newfoundland population-based CRC cohort. Probands that met the criteria to be in the NFCCR, that were not previously referred to the Provincial Medical Genetics Program were included in the cohort (n=481). Of the 481 DNA samples available, 478 (99.4%) were successfully sequenced for exons 4 and 6 of *GALNT12*. Approximately 47% of this cohort met the high or intermediate familial risk criteria outlined by Green and colleagues (2007) (Table 2.1).

In total, there were seven variants identified in 11 individuals (Table 4.5). Six variants were found in exon 4 and one variant was found in exon 6. Of these seven variants, two are considered to be putatively pathogenic while two are considered to be potentially non-pathogenic based on previous bioinformatic analysis and functional testing (Guda et al., 2009). For the remaining three variants, the effect the nucleotide changes have on the protein sequence is unknown. These variants are therefore considered to be unclassified variants (UCVs).

4.2.1 Putatively Pathogenic Variants

The putatively pathogenic variant, c.907G>A (p.Asp303Asn) already identified in three clinic-based families, was also identified in two population-based probands. This

Table 4.5 Germline variants in *GALNT12* identified in population-based probands (NFCCR).

Variant	Exon	Frequency		
		Cases	Controls	Population[1]
c.732-8G>T	Intron 3	0.002	0.000	ND
c.781G>A (p.Asp261Asn) (*rs41306504*)	4	0.006	0.012	0.008
c.857C>T (p.Thr297Met)	4	0.002	0.000	ND
c.868G>T (p.Val290Phe)	4	0.002	0.000	ND
c.889C>T (p.Arg297Trp)*	4	0.002	0.000	ND
c.907G>A (p.Asp303Asn)*	4	0.004	0.001	ND
c.1131G>A (p.=)	6	0.002	0.000	ND

[1]Data sourced from dbSNP [http://www.ncbi.nlm.nih.gov/projects/SNP/] via Ensembl [http://www.ensembl.org]; ND: not determined; * Indicates variants which are considered to be pathogenic based on previous functional testing (Guda et al., 2009).

variant was previously reported by Guda and colleagues (2009) to reduce enzymatic activity to 37% of WT function. Bioinformatic analysis predicts this variant, c.907G>A (p.Asp303Asn), to be damaging (see section 4.1.1).

The proband from family 1148 that was found to harbour this variant was diagnosed with CRC in the sigmoid colon at age 60 years (Figure 4.5). This individual had no family history of CRC but had an aunt who was diagnosed with stomach cancer at 80 years of age. The CRC tumour from the proband was found to be MSS in a previous study by Woods and colleagues (2010). They also tested the MMR proteins MLH1, MSH2 and MSH6 for deficiencies. These proteins were found to be intact in the tumour of the proband. No deleterious *MUTYH* variants were identified in the proband.

The same variant, c.907G>A (p.Asp303Asn), was also found in the proband from family 1220 (Figure 4.6) who was diagnosed with CRC in the descending colon at the age of 69 years. The proband's second cousin (PID 30) was diagnosed with CRC at the age of 51 years and a first cousin once removed (PID 27) was affected with CRC at age 60. There is a strong family history of several other types of cancer, particularly of the stomach. This proband does not exhibit any of the hallmark characteristics of LS as her tumour is MSS with intact MMR proteins (MLH1, MSH2, MSH6) (Woods et al., 2010). The proband also did not harbour any disease causing *MUTYH* variants. The family is considered low risk for CRC as it does not meet the intermediate or high risk criteria outlined by Green et al. (2007) (see Table 2.1).

The second putatively pathogenic variant found, c.889C>T (p.Arg297Trp), was identified in one individual from family 1543 (Figure 4.7). This family meets the

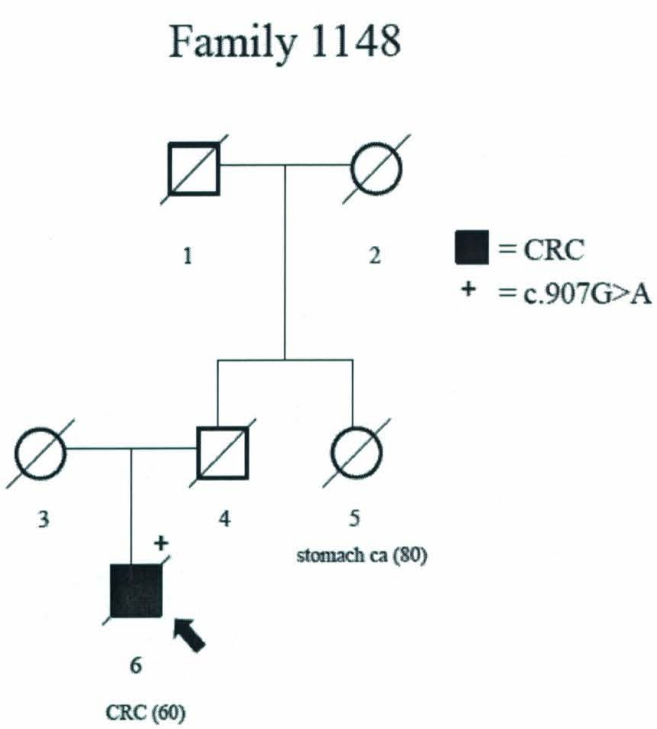

Figure 4.5 Family 1148. The proband harboured the c.907G>A, p. Asp303Asn *GALNT12* variant. No other family members were available for testing. Arrow indicates proband.

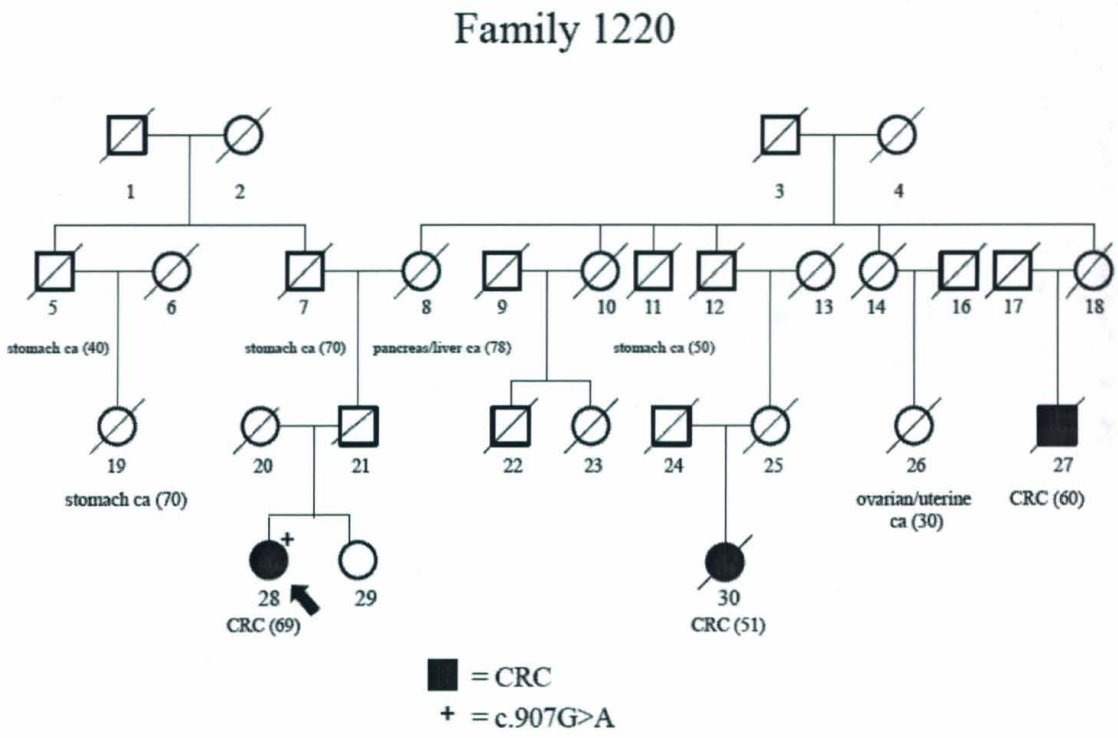

Figure 4.6 Partial pedigree of family 1220. The proband harboured the c.907G>A, p.Asp303Asn *GALNT12* variant. No other family members were available for testing. Arrow indicates proband.

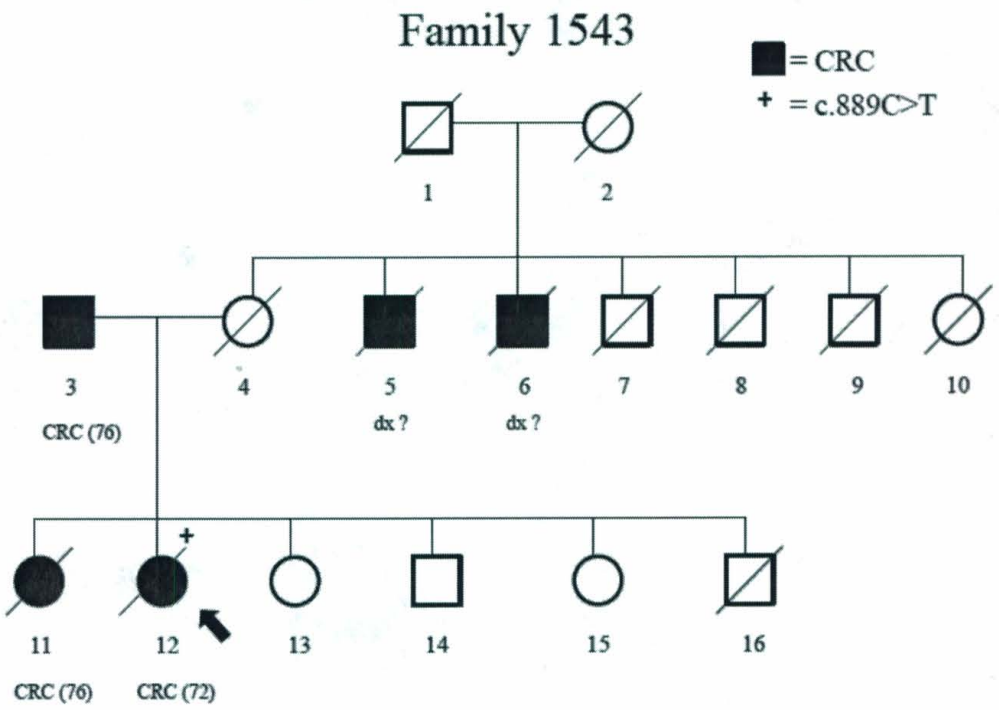

Figure 4.7 Partial pedigree for family 1543. The proband harboured the c.889G>A, p.Arg297Trp *GALNT12* variant. No other family members were available for testing. Arrow indicates proband.

intermediate risk-Revised Bethesda criteria; unfortunately no other family members were available for testing. This variant was previously reported in a CRC patient from the study conducted by Guda et al. (2009). They reported that this variant reduced the specific enzymatic activity to 7% of WT function. This variant was investigated using the bioinformatic tools PolyPhen, SIFT and Panther. PolyPhen predicts this variant as "probably damaging" with a score of 0.997 (a score of 0.0 being neutral and a score of 1.0 representing deleterious changes). SIFT predicts this change as "not tolerated" and Panther reported a subPSEC score of -7.54 and a $P_{deleterious}$ score of 0.99. The Align GVGD classification for this amino acid change, p.Arg297Trp, is C65 (Table 4.7). The proband harbouring this variant was diagnosed with CRC in the ascending colon at age 72 years. The proband's sister (PID 11) and father (PID 3) were also diagnosed with CRC, both at age 76 years. Two maternal uncles of the proband (PID 5, 6) were diagnosed with CRC as well (age unknown). The proband did not harbour any disease causing variants in the *MUTYH* gene. Additionally, the tumour of the proband was MSS and intact for MMR proteins (MLH1, MSH2 and MSH6) determined by IHC (Woods et al., 2010). The clinical information for individuals from the population-based CRC cohort harbouring a putatively pathogenic *GALNT12* variant in exon 4 or 6 is presented in Table 4.6.

4.2.2 Investigating Pathogenic Variants in Control Samples

Approximately 298 chromosomes from healthy individuals, with no family history of cancer, were sequenced in exons 4 and 6 to obtain the population frequency of the two potentially pathogenic variants, c.907G>A and c.889C>T. Neither variant was

Table 4.6 Putatively pathogenic *GALNT12* variants and population-based probands clinical information.

Study #	Variant	Proband Age of Diagnosis	Location of Cancer	MSI Testing[1]	IHC Testing[2]	Family Risk[3]
1148	c. 907G>A, p.Asp303Asn	60	Sigmoid Colon	MSS	intact	Low
1220	c.907G>A p.Asp303Asn	69	Descending Colon	MSS	intact	Low
1543	c.889C>T p.Arg297Trp	72	Ascending Colon	MSS	intact	B5

[1]MSS: microsatellite stable. [2]IHC testing was conducted on the MLH1, MSH2 and MSH6 proteins.
[3]Bethesda criteria: 1, CRC diagnosis at less than 50 years of age; 2, synchronous or metachronous tumours (colorectal or HNPCC); 4, one 1st degree relative with HNPCC related tumour under 50 years of age; 5, two 1st or 2nd degree relatives with HNPCC related tumours at any age. Family risk criteria are outlined by Green et al. (2007).

identified in the 149 healthy individuals tested from the NL population, p= 0.039 (Fisher's exact test).

4.2.3 Putatively Non-pathogenic Variants

Several variants that were considered to be putatively non-pathogenic were also identified by DNA sequencing. These variants were categorized as putatively non-pathogenic based on the population frequency of the variant greater than 1%, the variant previously reported in the control group in Guda et al. (2009), and/or the variant coded for a synonymous amino acid change.

The variant, c.781G>A (p.Asp261Asn), was identified in three population-based CRC cases (0.006). This variant has been reported in 0.8% of the population based on the 1000 genomes project as reported by dbSNP (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=41306504). Guda and colleagues (2009) also reported this variant in the control group with an enzymatic activity of 85% of WT function. Three of the 298 (1.0%) control sample chromosomes sequenced in this study had the c.781G>A variant. PolyPhen predicts p.Asp261Asn as "probably damaging" with a score of 0.997. SIFT predicts the change as "tolerated". Panther calculated a subPSEC score of -3.35 and a $P_{deleterious}$ of 0.586. The Align GVGD classification for this amino acid change, p.Asp261Asn, is C15 (Table 4.7).

A second putatively non pathogenic variant was identified, c.1131G>A (p. =). This variant was identified in one individual. This variant is considered as putatively non-pathogenic because it causes a synonymous change. This variant was not identified in the 298 chromosomes from healthy controls.

4.2.4 Unclassified Variants (UCV)

There were three novel variants identified within exon 4 of *GALNT12* that are considered to be of unknown pathogenicity (i.e. unclassified variants, UCVs). These variants were classified as such based on the limited information about the allele frequency in the general population and limited information about the effect they have on protein function.

The variant, c.732-8G>T, was identified in intron 3, eight bases before the splice site. The effect this variant has on the correct splicing of the *GALNT12* transcript is unknown. This variant was identified in one individual that was diagnosed with CRC of the transverse colon at 51 years. The proband belongs to family 1541 which meets the intermediate-risk Revised Bethesda criteria (Figure 4.8). The proband's maternal aunt and uncle (PID 6, 7) were both diagnosed with CRC at ages 37 and 82 years, respectfully. Additionally, the proband's maternal grandfather (PID 1) was also affected by CRC at age 65 years. This proband had an MSS tumour and the MMR proteins (MLH1, MSH2 and MSH6) were intact. The proband of this family does harbour a heterozygous, known pathogenic, variant in the *MUTYH* gene, p.Gly382Asp.

A second UCV was identified, c.857C>T (p.Thr286Met). This variant was identified in the proband from family 1595 who was diagnosed with CRC in the descending colon at age 60 years. This family meets the high-risk AC-1 criteria (see Figure 4.9). The proband's sister (PID 13) was diagnosed with "a cancer of the abdomen" at age 26 years; the proband's father (PID 8) was diagnosed with CRC at age 24 years. There was also a paternal uncle of the proband (PID 5) who was diagnosed with CRC

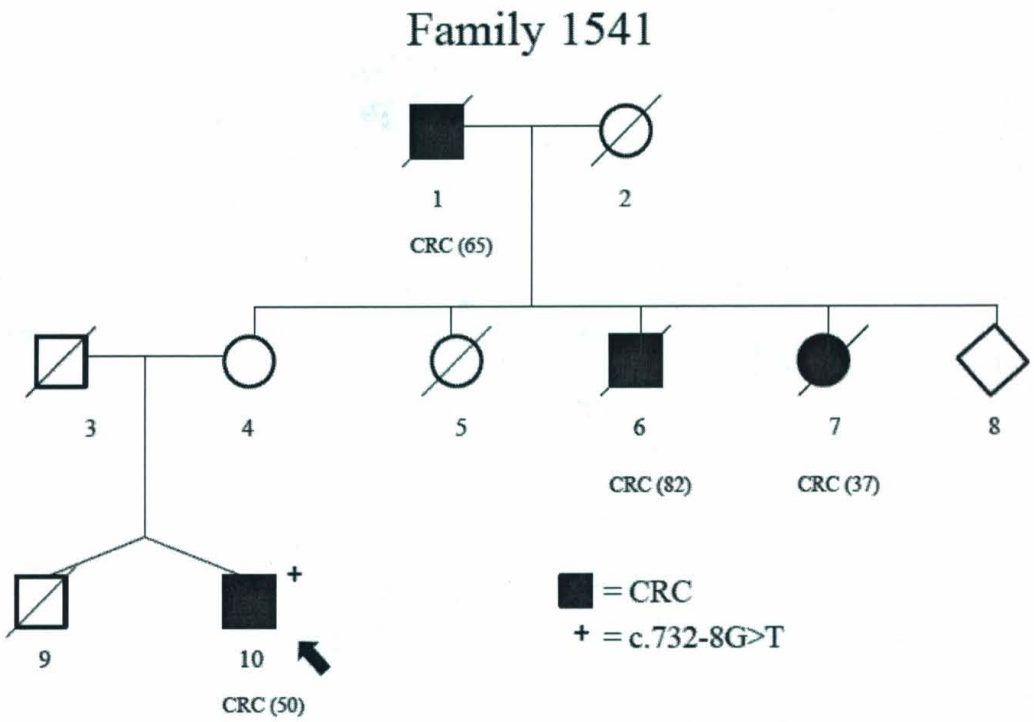

Figure 4.8 Partial pedigree for family 1541. The proband harbours the c.732-8G>T variant. The proband also harbours a known pathogenic variant in the *MUTYH* gene. No other family members were available for testing. Arrow indicates proband.

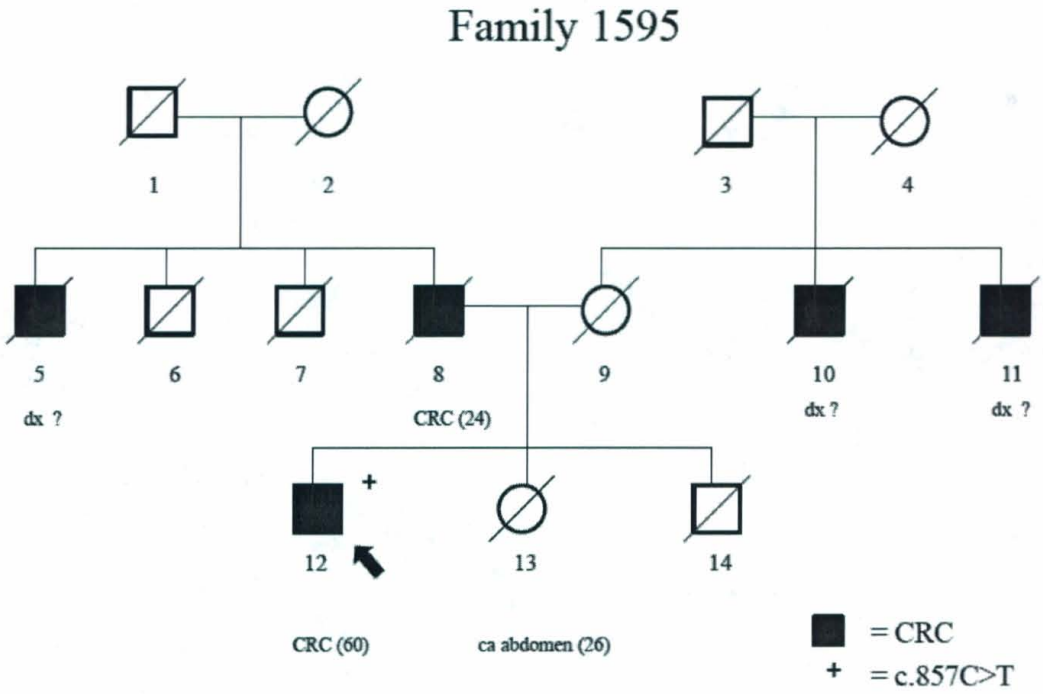

Figure 4.9 Partial pedigree for family 1595. The proband harboured the c.857C>T, p.Thr286Met *GALNT12* variant. No other family members were tested as this family has been diagnosed as having Lynch Syndrome-the proband harbours the pathogenic *MLH1* c793C>T variant. Arrow indicates proband.

(age unknown) and also two maternal uncles (PID 10, 11) diagnosed with CRC (age unknown). The tumour from this proband was found to be deficient in the MLH1 protein and was also found to have microsatellite instability (MSI). This individual harboured a pathogenic variant in *MLH1*, c.793C>T (Woods et al., 2010). The *GALNT12* variant, c.857C>T (p.Thr286Met), has not been previously reported in dbSNP. Bioinformatic analysis was conducted to determine the effect this variant has on protein function. PolyPhen predicts p.Thr286Met to be "probably damaging" with a score of 0.998. SIFT predicts the change to be "not tolerated" and Panther calculated a subPSEC score of -2.83 and a $P_{deleterious}$ of 0.456. The Align GVGD classification for this amino acid change, p.Thr286Met, is C65 (Table 4.7).

A third UCV, c.868G>T (p.Val290Phe), was identified in one individual. This proband is from family 2287 which is considered to be a low risk family (Figure 4.10). The individual harbouring this variant was diagnosed with CRC at age 64 years. The father of the proband (PID 1) was diagnosed with pancreatic cancer at the age of 81 years. The proband's brother (PID 5) was diagnosed with lung cancer at age 56 years. The proband's tumour is MSS and intact for the tested MMR proteins (MLH1, MSH2 and MSH6). The proband also does not harbour any disease causing variants in the *MUTYH* gene. Bioinformatic analysis was conducted to determine the effect this variant has on the function of the protein. PolyPhen predicts p.Val290Phe to be "probably damaging" with a score of 0.967. SIFT predicts the change to be "not tolerated" and Panther calculated a subPSEC score of -4.71 and a $P_{deleterious}$ of 0.847. The Align GVGD classification for this amino acid change, p.Val290Phe, is C45 (Table 4.7).

Table 4.7 Bioinformatic analysis of nonsynonymous *GALNT12* variants identified in population-based probands (NFCCR).

		Bioinformatic Tools			
Variant	Exon	SIFT[1]	PolyPhen[2] (Prediction, $P_{deleterious}$)	Panther[3] (subPSEC, $P_{deleterious}$)	Align GVGD[4]
c.781G>A (p.Asp261Asn)	4	TD	PrD, 0.997	-3.35, 0.586	C15
c.857C>T (p.Thr286Met)	4	NT	PrD, 0.998	-2.82, 0.456	C65
c.868G>T (p.Val290Phe)	4	NT	PrD, 0.967	-4.71, 0.847	C45
c.889C>T (p.Arg297Trp)	4	NT	PrD, 0.997	-7.54, 0.989	C65
c.907G>A (p.Asp303Asn)	4	NT	PrD, 0.999	-9.28, 0.998	C15

[1]TD: tolerated; NT: not tolerated. [2]PsD: possibly damaging; PrD: probably damaging. [3] subPSEC: the "negative logarithm of the probability ratio of the wild type and mutant amino acids at a particular position" (Brunham et al., 2005). These scores are continuous values that range from 0 (mostly likely benign) to -10 (most likely pathogenic).). [4]Classifications are listed from most likely to interfere with protein function to least likely; C65 (most likely), C55, C45, C35, C25, C15, C0 (less likely).

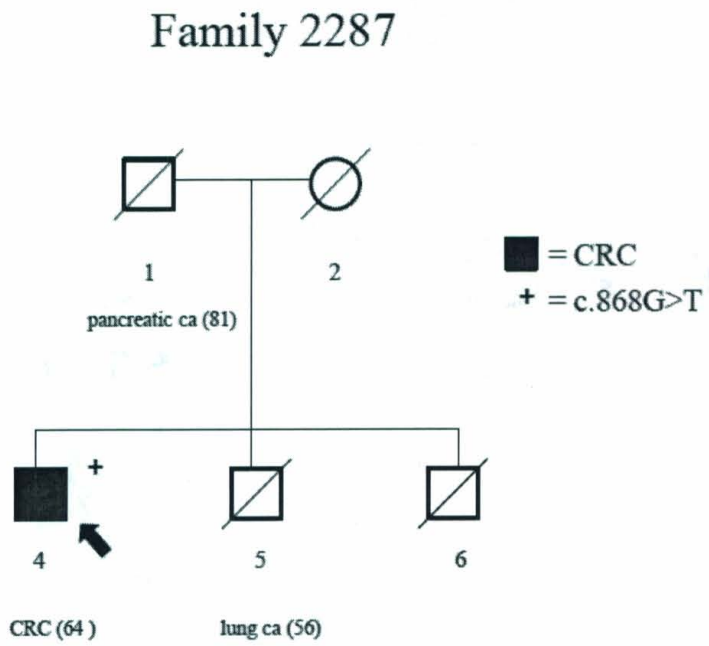

Figure 4.10 Partial pedigree for family 2287. The proband harboured the c.868G>T, p.Val290Phe *GALNT12* variant. No other family members were available for testing. Arrow indicates proband.

Chapter 5

5.0 Discussion

5.1 Clinic-based Cohort Summary

In total, three putatively pathogenic *GALNT12* variants were identified in five probands. Partial segregation of one of these variants with the disease phenotype was observed in two of the five families identified. A number of UCVs were identified which may also contribute to the onset of CRC. The results from the clinic-based portion of this study has been recently published (see appendix D, Clarke et al., 2012).

The clinic-based study identified two putatively pathogenic *GALNT12* variants, c.907G>A and c.1187A>G. The c.907G>A variant was identified in four probands and was found to segregate in two of the four families identified. This variant, c.907G>A, was previously reported in a study by Guda and colleagues (2009) to reduce enzymatic activity of the protein to 37% of WT function. Bioinformatic analysis also shows that this variant is predicted to be pathogenic (Table 4.8). Although the variant only partially segregated with the disease phenotype (some carriers not having the disease phenotype and some non-carriers diagnosed with CRC and/or adenomatous polyps), the SISA results revealed that the probability of co-segregation by chance, in both families (20444 and 20896), was 1.56%, indicating that this variant is likely associated with the disease phenotype observed. SISA is used to identify which lineages an unclassified variant is segregating in a hereditary cancer family. It also estimates the probability of chance co-segregation and can also be used to estimate the penetrance and/or expression.

Multiple individuals from family 20444 harbour the c.907G>A variant. When adenomatous polyps are included in the disease phenotype; four individuals (Figure 4.1) are affected in this family. Adenomatous polyps are considered to be an indication that an individual is at higher risk for developing CRC as these polyps have potential to progress to CRC.

All four affected individuals harbour the c.907G>A change. In addition, four other unaffected individuals (i.e. not diagnosed with CRC or adenomatous colonic polyps) also harboured the c.907G>A variant. It is worthy to note that the unaffected carriers are approximately 5-10 years younger than the affected carriers (unaffected carriers median age 54.5, range 50-60 years of age). It is possible that these individuals are at increased risk for developing CRC at a later age. It is also likely that the c.907G>A variant is not 100% penetrant. In summary, four affected and four unaffected individuals harbour the *GALNT12* variant, c.907G>A, and one individual with an unknown polyp type was negative for the variant. Through the use of SISA, we determined that the probability of co-segregation by chance with the disease phenotype was only 6.25%. These findings show support for the association between *GALNT12* and hereditary CRC in this family (20444).

Family 20896 also had multiple individuals that harboured the c.907G>A variant. If adenomatous polyps are included in the disease phenotype, a total of seven individuals are affected in this family. DNA was only available for four affected individuals as well as two individuals that were not affected. Three of the four affected individuals with available DNA carried the c.907G>A variant. One individual (PID18), who was affected with CRC at age 57, did not harbour the change. This individual was a son of an affected

mother (PID 7) who may or may not have carried the variant (DNA not available). It is possible that this affected non-carrier (PID 18) is a phenocopy (sporadic case) due to the high incidence of the disease in the overall population. Although finding an affected non-carrier somewhat weakens the argument for the involvement of *GALNT12* contributing to hereditary CRC within this family, three of the four affected individuals were carriers and two unaffected individuals were non-carriers. Additionally, the SISA analysis revealed a 25% probability that this variant co-segregated with the disease by chance. Overall, this family provides some evidence supporting the role of *GALNT12* in hereditary CRC.

Family 1728 provides little evidence for or against the association between *GALNT12* and CRC due to the fact that DNA was not available from other affected family members. The proband was young at the age of diagnosis (21 years) which is suggestive of genetic influences on CRC; however, the variant found was novel and has not been shown to be involved in with the disease. Bioinformatic analysis did, however, predict this variant to be pathogenic (Table 4.3). Functional testing should be conducted to define the effect this variant has on enzymatic function. More information is needed about the inheritance pattern of this variant within the family and the effect this variant has on protein function, to conclude if it is responsible for the presence of hereditary CRC in this family.

Four individuals from family 1117 were affected with CRC, three of whom had DNA available for testing. The proband carried the c.907G>A variant while the other two affected individuals, who are second cousins of the proband, did not (Figure 4.3). The possible involvement of *GALNT12* in the onset and progression of hereditary CRC in this family is difficult to interpret. It is possible that a different unknown genetic

predisposition is the underlying cause of CRC in this family, and the *GALNT12* variant identified in the proband is contributing exclusively to her observed phenotype.

The study by Guda and colleagues (2009) revealed findings that support the association between *GALNT12* and colon cancer. Several linkage and association studies have provided evidence for a gene responsible for CRC at the chromosomal region near *GALNT12* (9q22) (Wiesner et al., 2003; Kemp et al., 2006; Skoglund et al., 2006; Gray-McQuire et al., 2010). The present study is the first to use family-based studies to investigate the contribution of *GALNT12* to familial CRC. Two families that showed partial segregation of a variant, c.907G>A, with the disease phenotype. SISA revealed the probability by chance of co-segregation of c.907G>A with the disease phenotype, in both families, was 1.56%. This indicates that the variant, c.907G>A, occurring in these families is likely (98.44%) associated with the disease phenotype. The four individuals, from family 20444, that harbour this variant and have not had a history of CRC or adenomatous polyps can be explained by the late onset of the disease or incomplete penetrance of the variant. Guda and colleagues (2009) reported the average age of onset for CRC in cases identified with a causal *GALNT12* variant was 71.5 years. The four unaffected individuals identified with the c.907G>A variant in family 20444 had an average age of 54.5 years. It is possible that these individuals are at a higher risk for developing CRC later in life.

It is probable that if there was more DNA available from the family members included in the present study, the results may have been more conclusive in determining the impact *GALNT12* had on the disease phenotype. The results of this study were not as compelling as those from Guda et al., (2009) as they used functional testing to confirm

the presence of pathogenic variants and these pathogenic variants were only identified in cases and not controls. Although our results are not as convincing as those in Guda et al (2009), they are somewhat impactful as we did identify two putatively pathogenic variants in four affected individuals accounting for approximately 3.33% of the clinic-based study population.

Due to the novel association of *GALNT12* and CRC, the exact phenotype and types of cancers involved in this disease pathway remained to be determined. It is possible that *GALNT12* may be involved in other types of epithelial cancers. In the present study, the sister of the proband from family 20444 (PID 4; Figure 4.1) was affected with skin cancer at age 82. The sister of the proband from family 20896 (PID 8; Figure 4.2) was reported to have stomach cancer between the ages of 77 and 79 years (exact age unknown) and also had a subsequent mass in her transverse colon at age 82 years. The proband from family 1117 (PID18; Figure 4.3) who harboured the c.907G>A variant was also affected with skin cancer at age 74 years. Additionally, the paternal great uncle of the proband from family 1728 was diagnosed with stomach cancer at age 68 years. Furthermore, the study by Guda et al. (2009) reported having access to complete medical records from three cases that had an identified inactivating germline *GALNT12* variant. Each of these three individuals were reported to have been affected by two independent epithelial cancers which included independent synchronous CRC and multiple reports of breast cancer. It is possible that other extra-colonic epithelial cancers, such as stomach, breast and skin cancer are associated with the putative *GALNT12* pathway.

In summary, the results of the familial analysis provides partial evidence for the involvement of *GALNT12* in hereditary CRC; however, they do provide findings which strongly suggest that this gene should be further investigated to determine its impact on the onset and progression of this disease.

5.2 Population-Based Cohort Summary

A total of seven variants were identified by sequencing exons 4 and 6 in 478 individuals diagnosed with CRC. Exons 4 and 6 were selected for sequencing based on the findings from the clinic-based cohort. Two of the seven variants identified in the population-based probands have been considered to be putatively pathogenic based on previous findings (Guda et al., 2009).

The variant c.907G>A has been previously reported in CRC cases in the Guda study (2009). Through functional testing, Guda and colleagues (2009) found that this variant reduces enzymatic activity to 37% of WT function. Bioinformatic analysis also shows that this variant is predicted to be pathogenic (Table 4.8). This variant was the most frequent variant observed in both the population-based cohort and clinical-based cohort, being identified in a total of five CRC cases. This variant was not identified in any of the 149 healthy controls tested. It is worthy to note that 272 controls were initially sequenced in exons 4 and 6. One individual of the initial 272 controls harboured the c.907G>A variant. The control individual (age 65 years) that harboured this variant had a family history of epithelial cancer (Figure 4.11). Since mutations in *GALNT12* may cause others cancers too, all controls with a family history of cancer were removed from the study. As mentioned above, it may be possible that *GALNT12* pathogenic variants cause

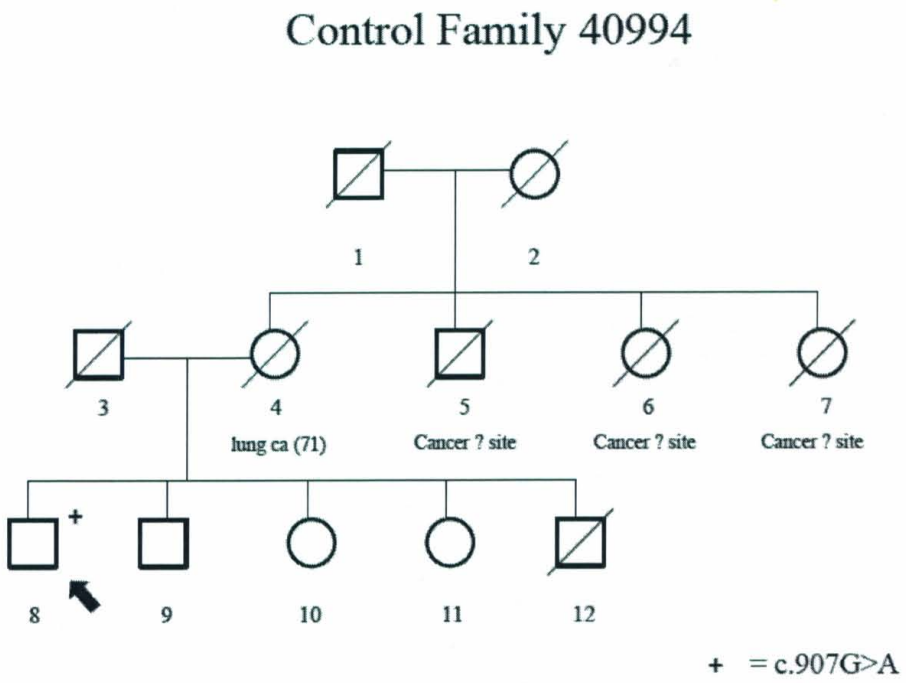

Figure 5.1 Partial pedigree for control family 40994. The control individual harboured the c.907G>A, p.Asp303Asn *GALNT12* variant. No other family members were available for testing. Arrow indicates control individual.

extra colonic carcinomas just as MMR pathogenic variants do in LS (Lynch et al., 2009). Therefore, this control individual may be at an increased risk for CRC or other carcinomas. Unfortunately DNA was not available from the family members of the population-based probands who harboured *GALNT12* variants. This DNA would have been useful in investigating the inheritance patterns of the variants identified, in addition to penetrance and variable expression patterns.

The second putatively pathogenic variant found in this cohort, c.889C>T, was identified in one individual. This variant has been previously reported by Guda et al. (2009) and functional testing revealed that this variant reduces enzymatic activity to 7% of the WT protein function. Bioinformatic analysis also predicted this variant to be pathogenic (Table 4.8). Based on Guda et al (2009), together with the bioinformatic analysis, this variant has been classified as putatively pathogenic. Unfortunately, DNA was not available from the family members of the proband to determine if the variant segregated with the disease phenotype.

There were several UCVs identified in the population based cohort. It was decided that bioinformatic analysis alone was not sufficient to classify a variant as pathogenic. More information is needed to conclude whether these variants significantly affect the function of the *GALNT12* protein and therefore all three UCVs should be further investigated by cDNA analysis and enzyme activity assays. cDNA analysis should be used to determine whether the c.732-8G>T variant, found in family 1541, affects splicing at the 3' end of intron 3. Enzymatic assays similar to those conducted by Guda et al. (2009) should be conducted for all three UCVs. It is possible that the c.857C>T variant contributed to the CRC phenotype observed in the proband from family 1541;

however, it is probable that this variant is not the main predisposition causing CRC in this patient as this individual also harboured a pathogenic *MLH1* variant, c.793C>T. The effect the c.857C>T *GALNT12* variant has on protein function should still be investigated to determine if this variant has an additive effect on the CRC phenotype observed in this family.

With advancements in technology providing increased opportunity for molecular genetic testing there has been a rapid increase in the identification of novel variants with unknown pathogenicity. Analysis of co-segregation of UCVs in high risk CRC families is a powerful approach to determine if the variant is pathogenic. Bioinformatic analysis based on amino acid conservation and physical properties is another approach used to determine the pathogenicity of UCVs. Case-control studies also provide frequencies of UCVs in both populations allowing for an estimation of whether the UCVs occur more in the disease population compared to the healthy population. Functional studies are also a useful approach to determine if a UCV is affecting the function of the protein. Specifically, the function of *GALNT12* can be measured through enzyme activity assays using transferase bound beads to measure the relative transferase activity compared to WT GALNT12 protein transferase activity (Guda et al., 2009). cDNA analysis is also another way to determine if the proper transcript is being transcribed when a potential splice-site variant is identified. While each of the above approaches have limitations, using these analyses together is an appropriate way to determine pathogenicity of UCVs.

In total, three cases out of the 478 included in the study were identified as harbouring putatively pathogenic variants accounting for 0.623% of the total CRC population studied. If these putatively pathogenic variants are truly pathogenic, the

proportion of CRC cases that the *GALNT12* variants account for in this study are comparable to that of FAP or MAP, which account for 1% and <1% of all CRC, respectively (Bisgaard et al., 1994; Croitoru et al, 2004; Fleischmann et al, 2004; Wang et al, 2004; Farrington et al, 2005). It may be possible that some of the UCVs identified are pathogenic, increasing the proportion of CRC cases accounted for by *GALNT12* causal variants. Additionally, it is likely that more individuals would have been identified with a *GALNT12* pathogenic variant if exons 1,2,3,5,7,8,9 and 10 were also investigated. Furthermore, it may be possible that screening for large exonic deletions and duplications through multiplex ligation-probe amplification (MLPA) analysis would uncover additional CRC cases.

5.3 GALNT12: The Potential Role in CRC

This study provides some evidence for the involvement of *GALNT12* in familial CRC. This data should encourage others to further investigate *GALNT12* variants in both clinical and population-based cohorts. If *GALNT12* is involved in the development of CRC it could be associated with the increase in improperly glycosylated or unglycosylated MUC1 proteins in the human colon. Guda and colleagues (2009) did find increased levels of unglycosylated MUC1 proteins in three out of three patients for which they had tumours available.

Mucin proteins are important for mucus production and providing a protective layer over the epithelium (Brockhausen, 2006). O-glycans make up over 50% of the total molecular weight of a mucin molecule. Glycosylation of mucin proteins is important for the maintenance of protein structure. It is possible that the protective function of the

mucin proteins may be compromised if the proper glycosylation of these proteins is not met. If pathogenic variants in *GALNT12* are associated with reduced glycosylation, it could be that this reduced glycosylation causes a reduced functional protection of the epithelial layer, therefore causing CRC to form in the epithelial cells that line the colon. Additionally, it has been noted that aberrant glycosylation can influence the growth and survival of a cell and its ability to metastasize (Brockhausen, 2006).

5.4 Study Limitations

There were a number of limitations in this study. First, both cohorts were investigated only for exonic point mutations. It is possible that a number of individuals included in the study may harbour large exonic deletions or duplications that would not have been identified through the methods used in this study. Second, cDNA analysis was not conducted to determine if the c.732-8G>T variant affected the correct splicing of the transcript. Third, in both cohorts, not all the affected family members of probands identified with a *GALNT12* variant had DNA available for testing. Not being able to test all affected individuals in the family limits the results and the ability to be conclusive about the segregation of the variant with the disease. Additionally, the validity of the bioinformatic tools used must be questioned as the algorithms sometimes give conflicting results. Bioinformatic tools should be used with caution and in conjunction with functional testing, segregation analysis or case-control analysis. Functional tests, such as enzyme activity assays used to deduce if the identified variants actually affected the enzymatic function of the *GALNT12* protein, were not performed in this study. Furthermore, IHC analysis investigating the proportion of properly glycosylated MUC1 protein within the colon tumours of individuals with *GALNT12* variants, as conducted by

Guda et al (2009), might have yielded more data to evaluate the variants identified in *GLANT12*. It is possible that glycosylation is not the only function of the GALNT12 protein. If this protein has other functions, variants in this gene could affect the additional functions of the protein. Thus enzymatic assays, employed by Guda et al., (2009) measuring glycosylation, may not have been effective in determining other functions of GALNT12 within the cell that could be contributing to the onset of CRC. Finally, any variants that were >30 bases into the intron were not assessed because they were not captured with the methods used. Such variants may have an effect on cryptic splice sites or other regulatory elements playing a role in expression and translation of this gene.

5.5 Future Studies

While the results of these two studies are not conclusive, they do suggest that further investigations into the role of *GALNT12* in CRC should be conducted. Future studies should include investigating all exons of *GALNT12* in population-based cohorts as well as determining if deletions and amplification mutations occur. Functional testing, cDNA analysis and studies that investigate large insertions, deletions and duplications should also be considered.

Further studies are needed to better understand the relationship between *GALNT12* and the onset and progression of CRC. Studies investigating *GALNT12* have the potential to provide data for knowledge translation from the laboratory to clinicians and patients, thus having a significant impact on survival, prevention and treatment of inherited CRC.

Chapter 6

6.0 Conclusion

In this study, two of four families harbouring a putatively pathogenic *GALNT12* variant had the variant segregate with CRC within their family. While there was only partial segregation of the variant with the disease phenotype, SISA revealed that the probability of co-segregation by chance was low, 1.56%. This study is the first to investigate the inheritance of *GALNT12* variants in hereditary CRC families.

The population-based study produced interesting results as it identified two pathogenic variants previously reported by Guda and colleagues (2009) in addition to identifying a number of UCVs that could also be important in the onset and progression of CRC.

Currently, there is a significant proportion of hereditary CRC that remains unsolved. Discovering a new genetic cause of hereditary CRC would add to the understanding of the etiology of this disease. While previous studies indicate an association between CRC and *GALNT12* (Guda et al., 2009; Gray-McQuire et al., 2010), the role of *GALNT12* in hereditary CRC remains unclear. Future studies should continue to investigate the inheritance of *GALNT12* variants to yield more conclusive results about this gene and its role in hereditary CRC.

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., et al. (2010). A method and server for predicting damaging missense mutations. *Nature Methods, 7*, 248-249.

Al-Tassan, N., Chmiel, NH., Maynard J., et al. (2002). Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nature Genetics, 30*, 227-232.

Ames, B.N. & Gold, L.S. (1991). Endogenous mutagens and the causes of aging and cancer. *Mutation Research, 250*, 3-16.

Amos, C.I., Keitheri-Cheteri, M.B., Sabripour, M., et al. (2004). Genotype-phenotype correlations in Peutz- Jeghers syndrome. *Journal of Medical Genetics, 41*, 327-333.

Baron, J.A., Cole, B., Sandler, R.S., et al. (2003). A randomized trial of aspirin to prevent colorectal adenomas. *New England Journal of Medicine, 348*, 891-899.

Bellinzzi, A.M., Frankel, W.L. (2009). Colorectal cancer due to deficiency in DNA mismatch repair function: A Review. *Advances in Anatomic Pathology, 1,* 405-417.

Benamouzig, R., Deyra, J., Martin, A ., et al. (2003). Daily soluble aspirin and prevention of colorectal adenoma recurrence: one-year results of the APACC trial. *Gastroenterology, 125*, 328-336.

Bisgaard, M.L., Fenger, K., Bulow, S., Niebuhr, E. & Mohr J. (1994). Familial adenomatous polyposis (FAP): Frequency, penetrance, and mutation rate. *Human Mutation, 3*, 121-125.

Bond, J.H. (2000). Polyp Guidline: Diagnosis, treatment and surveillance for patients with colorectal polyps. *American Journal of Gastroenterology, 95*, 3053-3063.

Brockhousen, I. (2006). Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Reports, 7,* 599-604.

Brunham, L.R., Singaraja, R.R., Pap, T.D., Kejariwal, A., Thomas, P.D., Hayden, M.R. (2005). Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the *ABCA1* gene. *PLoS Genetics, 1*, 739-747.

Burn, J., Bishop, T., Mecklin, J.P., et al. (2008). Effect of aspirin or resistant starch on colorectal neoplasia in the Lynch syndrome. *New England Journal of Medicine, 359*, 2567-2578.

Burn, J., Gerdes, A.M., Macrae, E., et al. (2011). Long term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomized control trial. *Lancet* (epub ahead of print).

Bussey, H.J.R. (1975). *Familial polyposis coli*. Baltimore: John Hopkins University Press.

Byrd, J.C. & Bresalier, R.S. (2004) Mucins and mucin binding proteins in colorectal cancer. *Cancer and Metastasis Review, 23*, 77-99.

Canadian Cancer Society's Steering Committee: Canadian Cancer Statistics 2011, Toronto: Canadian Cancer Society 2010. April 2010, ISSN 0835-2976.

Canadian Cancer Society (2011) *Colorectal cancer statistics at a glance*. Retrieved from http://www.cancer.ca/canada-wide/about%20cancer/cancer%20statistics/stats%20at%20a%20glance/colorectal%20cancer.aspx

Canadian Partnership Against Cancer (n.d) *Colorectal cancer incidence and mortality*. Retrieved from www.cancerview.ca

Cheadle, J.P. & Sampson, J.R. (2007). *MUTYH*-associated polyposis-from defect in base excision repair to clinical genetic testing. *DNA Repair, 6*, 274-279.

Chen, J., Giovannucci, E.L., Hunter, D.J. (1999). MTHFR polymorphism, methyl-replete diets and the risk of colorectal carcinoma and adenoma among U.S. men and women: an example of gene-environment interactions in colorectal tumorigenesis. *Journal of Nutrition, 129*, 560S-564S.

Clarke, E., Green, R.C., Green, J.S., Mahoney, K., Parfrey, P.S., Younghusband, H.B., Woods, M.O. (2012). Inherited deleterious variants in *GALNT12* are associated with CRC susceptibility. *Human Mutation, 00*, 1-3.

Croitoru, M.E., Cleary, S.P., Di Nicola, N., et al. (2004). Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. *Journal of the National Cancer Institute,* 96,1631-1634.

De Vos, M., Hayward, B.E., Picton, S., Sheridan, E., Bonthron, D.T. (2004). Novel PMS2 pseudogenes can conceal recessive mutations cause a distinctive childhood cancer syndrome. *American Journal of Human Genetics, 74,* 954-964.

DiSario, J.A., Foutch, P.G., Mai, H.D. et al. (1991). Prevalence and malignant potential of colorectal polyps in asymptomatic, average-risk men. *American Journal Gastroenterology, 86*, 941-945.

Doll, R. & Peto, R. (1981) The causes of cancer quantitative estimates of avoidable risks of cancer in the United States today. *Journal of the National Cancer Institute, 66*,1191-1308.

Dove-Edwion, I., De Jong, A.E., Adams, J., et al. (2006). Prospective results of surveillance colonoscopy in dominant familial colorectal cancer with and without Lynch syndrome. *Gastroenterology, 130,*1995-2000.

Entius, M.M., Keller, J.J., Drillenburg, P., Kuypers, K.C., Giardiello, F.M., Offerhaus, G.J. (2000). Microsatellite instability and expression of hMLH-1 and hMSH-2 in sebaccous gland carcinomas as markers for Muir-Torre syndrome. *Clinical Cancer Research, 6*, 1784-1789.

Farrington, S.M., Tenesa, A., Barnetson, R., et al. (2005). Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *American Journal of Human Genetics, 77,*112-119.

Felton, K.E.A., Gilchrist, D.M. & Andrew, S.E. (2007). Constitutive deficiency in DNA mismatch repair: is it time for Lynch III? *Clinical Genetics, 71,* 499-500.

Fleischmann, C., Peto, J., Cheadle, J., et al. (2004). Comprehensive analysis of the contribution of germline MYH variation to early-onset colorectal cancer. *International Journal of Cancer, 109*, 554-558.

Freeman, H.J. (2008). Colorectal cancer risk in Crohn's disease. *World Journal of Gastroenterology, 14*, 1810-1811.

Friedman, C.J. & Fechner, R.E. (1982). A solitary juvenile polyp with hyperplastic and adenomatous glands. *Digestive Diseases of Sciences, 27*, 946-948.

Frio, T.R., Lavoie, J., Jamel, N., et al. (2010). Homozygous *BUB1B* mutation and susceptibility to gastrointestinal neoplasia. *New England Journal of Medicine, 363,* 2628-2637.

Garcia-Castillo, H., Vasquez-Velasquez, A.I., Rivera, H., Barros-Nunez, P. (2008). Clinical and genetic heterogeneity in patients with mosaic variegated aneuploidy: delineation of clinical subtypes. *American Journal of Medical Genetics, 146,*1687-1695.

Giardiello, F.M.& Trimbath, J.D. (2006). Peutz-Jeghers syndrome and management recommendations. *Clinical Gastroenterology and Hepatology, 4*, 408-415.

Giovannucci, E., Stampfer, M.J., Colditz, G., et al. (1992). Relationship of diet to risk of colorectal adenoma in men. *Journal of the National Cancer Institute, 84*, 91-98.

Giovannucci, E, Stampfer, M.J., Colditz, G., et al. (1993). Folate, methionine, and alcohol intake and risk of colorectal adenoma. *Journal of the National Cancer Institute, 85,* 875-884.

Goss, K.H. & Groden, J. (2000). Biology of the adenomatous polyposis coli tumour suppressor. *Journal of Clinical Oncology, 18*, 1967-1979.

Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science, 185,* 862-864.

Grau, M.V., Baron, J.A., Sandler, R.S., et al. (2003). Vitamin D: Calcium supplementation, and colorectal adenomas: Results of a randomized trial. *Journal of the National Cancer Institute, 96*, 1765-1771.

Gray-McGuire, C., Guda, K., Adrianto, I., et al. (2010). Confirmation of linkage to and localization of familial colon cancer risk haplotype on chromosome 9q22. *Cancer Research, 70*, 5409-5418.

Green, R.C., Green, J.S., Buchler, S.K., et al. (2007). Very high incidence of familial colorectal cancer in Newfoundland: a comparison with Ontario and 13 other population-based studies. *Familial Cancer, 6*, 53-62.

Green, J., O'Driscoll, M., Barnes, A., et al. (2002). Impact of gender and parent of origin on the phenotypic expression of hereditary nonpolyposis colorectal cancer in a large Newfoundland kindred with a common MSH2 mutation. *Diseases of the Colon and Rectum, 45*, 1223-1232.

Guda, K., Moinova, H., He, J., et al. (2009). Inactivating germ-line and somatic mutations in polypeptide *N*-acetylgalactosaminyltransferase 21 in human colon cancers. *Proceedings of the National Academy of Science, 106,*12921-12925.

Guo, J.M., Zhanga, Y., Chenga, L., et al. (2002). Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases family, pp-GalNac-T12. *FEBS Letters, 524,* 211-218.

Hagen, F.K., Hazes, B., Raffo R., deSa, D., Tabak, L. (1999). Structure-function analysis of the UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase. *Journal of Biological Chemistry, 274,* 6797-6803.

Haidle, J.L. & Howe, J.R. (2003). Juvenile Polyposis Coli Includes: BMPR1A-Related Juvenile Polyposis; SMAD4-related Juvenile Polyposis. In :Pagon, R.A., Bird, T.D., Dolan, C.R. et al. *Juvenile Polyposis* Seattle (WA): GeneReviews.

Hamilton, S.R., Liu, B., Parsons, R.E., et al. (1995). The molecular basis of Turcot's Syndrome. *New England Journal of Medicine, 332,* 839-847

Hashimoto, R., Fujitani, N., Takegawa, Y., et al. (2011). An efficient approach for the characterization of mucin-type glycopeptides: The effect of O-glycosylation on the conformation of synthetic mucin peptides. *Chemistry - A European Journal,* Article in press.

Howe, J.R., Roth, S., Ringold, J.C., et al. (1998b). Mutations in the SMAD4/DPC4 gene in juvenile polyposis. *Science, 280,* 1086-1088.

Hsieh, P. & Yamane, K. (2008). DNA mismatch repair: Molecular mechanism, cancer, and aging. *Mechanisms of Ageing and Development, 129*, 391-407.

Hyman, N. H., Anderson, P. & Blasyk, H. (2004). Hyperplastic polyposis and the risk of colorectal cancer. *Diseases of the Colon and Rectum, 47,* 2101-2104.

Iyer, R.R., Pluciennik, A., Burdett, V. & Modrich, P.L. (2006). DNA mismatch repair: Functions and mechanisms *Chemical Reviews, 106,* 302-323.

Jarvis, L., Bathurst, N., Mahan, D. & Beckly, D. (1988). Turcot's syndrome: A review. *Diseases of the Colon and Rectum, 31,* 907-914.

Jass, J.R. (2006). Hereditary non-polyposis colorectal cancer: the rise and fall of a confusing term. *World Journal of Gastroenterology, 12*, 4943-4950.

Jass, J.R. (2001). Serrated route to colorectal cancer: back street or super highway? *Journal of Pathology, 193,* 283-285.

Jenkins, M.A., Bagietto, L., Dowty, J.G, et al. (2006). Cancer risks for mismatch repair gene mutation carriers: A population-based early onset case-family study. *Clinical Gastroenterology and Hepatology, 4,* 489-498.

Kaplan, K.B., Burds, A.A., Swedlow, J.R., Bekir, S.S., Sorger, P.K., Nathke, I.S. (2001). A role for the Adenomatous Polyposis Coli protein in chromosome segregation. *Nature Cell Biology, 3*, 429-432.

Kemp, Z.E., Carvajal-Carmona, L.G., Barclay, E., et al. (2006). Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. *Cancer Research, 66,* 5003-5006.

Kemp, Z., Thirlwell, C., Sieber, O., Silver, A., Tomlinson, I. (2004). An update on the genetics of colorectal cancer. *Human Molecular Genetics, 13,* R177-R185.

Kinzler, K.W., Nilbert, M.C., Su L-K, et al. (1991). Identification of FAP locus gene from chromosome 5q21. *Science, 253*, 661-665.

Kinzler, K.W., Volgelstein, B. (2002). Colorectal tumors. In: Vogelstein B, Kinzler KW (Ed.) *The genetic basis of human cancer.* New York: McGraw Hill

Knudson, A.G. (1971). Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Science, 68*, 820-823.

Kolodner, R. (1996). Biochemistry and genetics of eukaryotic mismatch repair. *Genes and Development, 10,* 1433-1442.

Le Marchard, L., Donlon, T., Hankin, J.H., Kolonel, L.N., Wilkens LR & Seifried A. (2002). B-vitamin intake, metabolic genes, and colorectal cancer risk (United States). *Cancer Causes Control, 13*, 239-248.

Lindor, N.M. (2009). Familial Colorectal Cancer Type X. In: Potter, J.D. & Lindor, N.M. (eds.) *Genetics of Colorectal Cancer*: Springer Science.

Logan, R.F.A., Grainge, M.J., Shepherd, V.C., Armitage, N.C. & Muir, KR. (2008). Aspirin and folic acid for the prevention of recurrent colorectal adenomas. *Gastroenterology, 134,* 29-38.

Lynch, H.T., Lynch, J.F.& Attard, T.A. (2009). Diagnosis and management of hereditary colorectal cancer syndromes: Lynch syndrome as a model. *Canadian Medical Association Journal, 181,* 273-280.

Lynch, H.T., Lynch, P.M., Lanspa, S.J., et al. (2009). Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clinical Genetics. 76*, 1-18.

Lynch, H.T., de la Chapelle, A. (2003). Genomic medicine: hereditary colorectal cancer *New England Journal of Medicine, 348*, 919-932.

Mannion, J.J. (1977). *The Peopling of Newfoundland: Essays in Historic Geography*. St. John's, Memorial University of Newfoundland.

Markowitz, A.J. & Winawer, S.J. (1997). Management of colorectal polyps. *Cancer Journal for Clinicians, 47*, 93-112.

Mathe, E., Olivier, M., Kato, S., Ishioka, C., Hainaut, P. & Tavtigian, S.V. (2006). Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Research, 34,* 1317-1325.

Miyoshi, Y., Ando, H., Nagse, H., et al. (1992). Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. *Proceedings of the National Academy of Science, 89,* 4452-4456.

Møller, P., Clark, N. & Mæhle, L. (2011). A simplified method for segregation analysis (SISA) to determine penetrance and expression of a genetic variant in a family *Human Mutation, 32,* 568-571.

Mueller-Koch, Y., Volgelsang, H., Koop R., et al. (2005). Hereditary non-polyposis colorectal cancer: clinical and molecular evidence for a new entity of hereditary colorectal cancer. *Gut, 54,* 1733-1740.

Ng, P.C. & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Research, 11,* 863-874

Nieminen, T. T., Abdel-Ragman, W. M., Ristmaki, A., Lappalainen, M., Mechklin, J.P. & Jarvinen, H. J. (2011). BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency. *Gastroenterology, 141,* e23-e26.

Nieuwenhuis, M.H. & Vasen, H.F.A. (2007). Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): A review of the literature. *Critical Reviews in Oncology/Hematology, 61,* 153-161.

O'Brien, M.J., Winawer, S.J., Zauber, A.G., et al. (1990). The National Polyp study: Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. *Gastroenterology, 98,* 371-379.

Olufemi, S.E., Green, J.S., Manickam, P., et al. (1998). Common ancestral mutation in the MEN1 gene is likely responsible for the prolactinoma variant of MEN1 (MEN1Brin) in four kindreds from Newfoundland. *Human Mutation, 11,* 264-269.

Prieto, V.G. (2011) *Muir-Torre Syndrome.* Retrieved from http://emedicine.medscape.com/article/1093640-overview

Rahman, P., Jones, A., Curtis, J., et al. (2003). The Newfoundland population: a unique resource for genetic investigation of complex diseases. *Human Molecular Genetics, 12,* R167-R172.

Rickert, R.R., Auerbach, O., Garfinkel, L., et al. (1979). Adenomatous lesions of the large bowel: An autopsy study. *Cancer,* 43, 1847-1857.

Rodriguez-Bigas, M.A., Boland, C.R., Hamilton, S.R., et al. (1997). A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer

Syndrome: meeting highlights and Bethesda guidelines. *Journal of the National Cancer Institute, 89,* 1758-1762.

Rothwell, P.M., Wilson, M., Elwin, C.E., et al. (2010). Long-term effect of aspirin on colorectal cancer incidence and mortality: 20 year follow-up of five randomized trials. *The Lancet, 376,* 1741-1750.

Ryan-Harshman, M. & Aldoori, W. (2007). Diet and colorectal cancer. *Canadian Family Physician, 53,* 1913-1920.

Sampson, J.R. & Jones, N. (2009). *MUTYH*-associated polyposis. *Best Practice and Research Clinical Gastroenterology, 23,* 209-218.

Sandler, R.S., Halabi, S., Baron, J.A., et al. (2003). A randomized trial of aspirin to prevent colorectal adenomas in patients with previous colorectal cancer. *New England Journal of Medicine, 348*, 883-890.

Sieber, O.M., Segditsas, S., Knudsen, A., et al. (2006). Disease severity and genetic pathways in attenuated familial adenomatous polyposis vary greatly, but depend on the site of the germline mutation. *Gut,* 55,1440-1448.

Sieber, O.M., Tomlinson, I.P. & Lamlun, H. (2000). The adenomatous polyposis coli (*APC*) tumour suppressor-genetics, function and disease. *Molecular Medicine Today,* 6, 462- 469.

Skoglund, J., Djureinovic, T., Zhou, X.L., et al. (2006). Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1. *Journal of Medical Genetics* 43;e7 Retrieved from http://www.jmedgenet.com/cgi/content/full/43/2/e7

Snover, D.C., Jass, J.R., Fenoglio-Preiser, C. & Batts, K. P. (2005). Serrated polyps of the large intestine: a morphologic and molecular review of an evolving concept. *American Journal of Clinical Pathology, 124,* 380-391.

Spirio, L, Green, J., Robertson, J. et al. (1999). The identical 5' splice-site acceptor mutation in five attenuated APC families from Newfoundland demonstrates a founder effect. *Human Genetics, 105*, 388-398.

Statistics Canada (2010) Retrieved from http://www.statcan.gc.ca/daily-quotidien/101222/dq101222a-eng.htm

Tavtigian S.V., Deffenbough, A.M., Yin, L., et al. (2005). Comprehensive statistical study of 452 BRCA1 missense substitutions with classifications of eight recurrent substitutions as neutral. *Journal of Medical Genetics,* epub ahead of print.

Ten Hagen, K.G., Fritz, T.A. & Tabak, L.A. (2003). All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology, 13*,1R-16R.

Thomas, P.D., Kejariwal, A. (2004). Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: Evolutionary evidence for differences in molecular effects. *Proceedings of the National Academy of Science, 101*,15398-15403

Torlakovic, E., Snover, D.C. (1996). Serrated adenomatous polyposis in humans. *Gastroenterology, 110,* 748-755.

Tung-Hua, L., Min-Chang, C., Hsien-Chiu, L.C,. et al. (1978). Malignant change of juvenile polyp of the colon: A case report. *Chinese Medical Journal (English), 4,* 434-439,

Valle, L., Perea, J., Carbonell, P., et al. (2007). Clinicopathologic and pedigree differences in Amsterdam-I-positive hereditary nonpolyposis colorectal cancer families according to tumor microsatellite instability status. *Journal of Clinical Oncology, 25*, 781-786

Vatn, M.H. & Stalsberg, H. (1982). The prevalence of polyps of the large intestine in Oslo: An autopsy study. *Cancer, 49*, 819-825.

Vasen, H.F.A. (2000). Clinical diagnosis and management of hereditary colorectal cancer syndromes. *Journal of Clinical Oncology,* 18, 81s-92s.

Wang, L., Baudhuin, L.M., Boardman, L.A. et al. (2004). MYH mutations in patients with attenuated and classic polyposis and with young- onset colorectal cancer without polyps. *Gastroenterology, 127*, 9-16.

Waye, J.D., Lewis, B.S., Frankel, A.et al. (1988). Small colon polyps. *American Journal of Gastroenterology, 83*, 120-122.

Willett, W.C., Stampfer, M.J., Colditz, G.A., Rosner, B.A. & Speizer, F.E. (1990). Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women. *New England Journal of Medicine, 323*,1664-1672

Wiesner, G.L., Daley, D., Lewis, S., et al. (2003). A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. *Proceedings of the National Academy of Science, 100*,12961-12965.

Woods, M.O., Hyde, A.J., Curtis, F.K., et al. (2005). High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. *Clinical Cancer Research, 11,* 6853-6861.

Woods, M.O., Williams, P., Careen, A., et al. (2007). A new variant database for mismatch repair genes associated with Lynch syndrome. *Human Mutation, 28*, 669-673.

Woods, M.O., Younghusband, H.B., Parfrey, P.S., et al. (2010). The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. *Gut, 59*, 1369-1377.

Xie, Y.G., Zheng, H., Leggo, J., et al. (2002). A founder factor VIII mutation, valine 2016 to alanine, in a population with an extraordinarily high prevalence of mild hemophilia A. *Journal of Thrombosis and Haemostasis*, *87*,178-179.

Young, T.L., Penney, L., Woods, M.O., et al. (1999). A fifth locus for Bardet-Biedl syndrome maps to chromosome 2q31. *American Journal of Human Genetics, 64*, 900-904.

Appendix A: Amsterdam Criteria

Amsterdam Criteria I (AC-I)

- Three relatives with CRC plus all of the following:
  - One affected 1st degree of other 2
  - Two successive generations affected
  - One diagnosis at < 50 years
  - FAP excluded
  - Tumours verified by pathological examination

Amsterdam Criteria II (AC-II)

- Three relatives with Lynch Syndrome[1] related cancer and all of the following:
  - One affected first degree relative of the other 2
  - Two successive generations affected
  - One diagnosis at age <50 yrs
  - FAP excluded
  - Tumours verified by pathological examination

[1]Lynch Syndrome related cancers include; endometrial/uterine, small bowel, kidney and ureter

Appendix B: Revised Bethesda Criteria

Tumors from individuals should be tested for MSI in the following situations:

1. Colorectal cancer diagnosed in a patient who is less than 50 years of age.

2. Presence of synchronous, metachronous colorectal, or other LS associated tumors[1], regardless of age.

3. Colorectal cancer with the MSI-H histology diagnosed in a patient who is less than 60 years of age.

4. Colorectal cancer diagnosed in one or more first-degree relatives with an LS-related tumor[1], with one of the cancers being diagnosed under age 50 years.

5. Colorectal cancer diagnosed in two or more first- or second-degree relatives with LS-related tumors[1], regardless of age.

[1]Lynch Syndrome related cancers include; endometrial/uterine, ovarian, small bowel, pancreatic, kidney, ureter, hepato-biliary, brain and lymphoma

Appendix C: *GALNT12* Primer Sequences

Exon	Forward Primer	Reverse Primer
1	5'-CTGCCGAAATGAGCCAC-3'	5'-GCGGTCGCTGAGGTAGATG-3'
2	5'- GTCCCCTTTGTCACTCCATC-3'	5'-CTGGGCCTACTTTAGCCCTC -3'
3	5'- ACAGGGAGGCCCAGAAGG-3'	5'-TGATGAGTCTGTTCCTCATGC-3'
4	5'- TGAATTTCCCAATTGTCTTCC-3'	5'-ACAGCTAACCATTCCTCCCG-3'
5	5'- TCGTGGTGGGAAGGTAGTTG-3'	5'-GGTCCACCAGGCTCTTACAG-3'
6	5'- GGCCTCTTTGCTGCAGATAC -3'	5'-CCAGATCGACTATGTTGGGG-3'
7	5'- CAGTGAACACAGGGGCTTTG-3'	5'-CATGACGTGGCCTTTCTTG-3'
8	5'- TCTTGAAGCAGGCTCTCCTC -3'	5'-TGGCCACCCTAGGCTTCTAC -3'
9	5'- TTGTCCAGCGATCTTTCCTC-3'	5'-CCTCACCAGGACTATGCCAG-3'
10a	5'- CTGCGTTACACGGAAGACACT-3'	5'-TACGATTACCCTTTGAAAATAGG-3'
10b	5'- AAACTAGGCTGCATTGCTTTG-3'	5'-GGAAAACTTTAACCCCGCAG-3'
10c	5'- CAAGAATTCCCAGGTACGAAG-3'	5'CATTTCTCCCTGACACTACCG-3'
10d	5'- GAGAACAATTTGCTTTACTAAGCTG -3'	5'-CAAGTGATCTGCCCTCCTTG-3'

Appendix D: Inherited Deleterious Variants in *GALNT12* are Associated with CRC Susceptibility

See attached document.

BRIEF REPORT

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Inherited Deleterious Variants in GALNT12 are Associated with CRC Susceptibility

Erica Clarke,[1] Roger C. Green,[1] Jane S. Green,[1] Krista Mahoney,[1] Patrick S. Parfrey,[2] H. Banfield Younghusband,[1] and Michael O. Woods[1]*

[1]Discipline of Genetics, Memorial University, St. John's, NL, Canada; [2]Clinical Epidemiology Unit, Memorial University, St. John's, NL, Canada

Communicated by Stephen J. Chanock

Received 23 December 2011; accepted revised manuscript 9 March 2012.
Published online 27 March 2012 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.22088

ABSTRACT: A recent report detailed the occurrence of both somatic and constitutional variants in the *GALNT12* gene, located at 9q22.33, in some colorectal cancer (CRC) patients. In this study, we investigate the occurrence of inherited deleterious variants in *GALNT12* in 118 families referred to a cancer genetics clinic. We discovered two deleterious variants (c.907G>A (p.Asp303Asn); c.1187A>G (p.Tyr396Cys)) in 4/118 probands. The variants, which were not found in 149 control individuals ($P = 0.0376$), cosegregate with CRC and/or adenomatous polyps in other family members. The probability by chance that cosegregation of c.907G>A with CRC and/or adenomatous polyps occurred, in the two pedigrees combined, was 1.56%. Although this study does not provide irrefutable evidence that *GALNT12* variants are highly penetrant alleles that predispose to CRC in the majority of unexplained hereditary CRC families, it does provide additional evidence to support an important role of these variants in a proportion of this considerable high-risk group.

Hum Mutat 00:1–3, 2012. © 2012 Wiley Periodicals, Inc.

KEY WORDS: GALNT12; hereditary colorectal cancer; CRC; Newfoundland

In about 5% of colorectal cancer (CRC) patients, a strong family history can be explained by highly penetrant alleles occurring in any of several well-characterized genes but, in many other such families, the genetic basis remains obscure.

The glycoproteins of colorectal tumors are often qualitatively or quantitatively abnormal. In a 2009 study of colorectal tumors, Guda et al. [2009] described finding deleterious variants in a gene coding for an enzyme involved in the O-glycosylation of mucin-type glycans. This enzyme, *N*-acetylgalactosaminyltransferase-type 12 (GALNT12), is expressed at high level in the normal colon [Guo et al., 2002]. Guda et al. [2009] found seven deleterious germline variants in *GALNT12* (MIM# 610290) in colorectal cancer (CRC) patients, which were not found in a control group. However, no information was available on the family history of the CRC patients carrying these germline *GALNT12* variants. We now report on the occurrence of *GALNT12* variants in a cohort of CRC families referred to a cancer genetics clinic.

Newfoundland and Labrador (NL) has the highest incidence of CRC in Canada [Canadian Cancer Statistics, 2011], being 49% higher than the Canadian average. Investigations of a large cohort of consecutive NL CRC patients revealed a high frequency of familial CRC that may be attributable to as yet undiscovered susceptibility genes [Green et al., 2007; Woods et al., 2010]. Given this high incidence of unexplained familial CRC, we decided to determine the occurrence of *GALNT12* variants in the probands of 118 CRC families, which had been referred to a genetics clinic because of an extensive family history of CRC and/or CRC diagnosis at a young age.

Approximately 22% ($n = 26$) of the probands met the Amsterdam I criteria and the remaining 78% ($n = 92$) met the revised Bethesda criteria [Umar et al., 2004; Vasen et al., 1991]. Family members gave informed consent to be included in CRC studies and provided a blood sample and access to tissue samples and medical records. Ethical approval was obtained from Memorial University's Human Investigation Committee. Primer sequences for each of the 10 exons of *GALNT12* were described by Guda et al. [2009]. All exons and intron/exon boundaries were screened for variants using standard protocols for automated direct sequencing. Primer sequences and PCR conditions are available upon request. PCR products were sequenced on an ABI Sequencer 3130XL and data was analyzed using Sequencing Analysis 5.2 and Sequencher 4.9. All variants identified are described in Supp. Table S1.

In total, two pathogenic variants were identified in probands from four different families (Table 1 and Supp. Table S1). Three individuals harbored a c.907G>A (p.Asp303Asn) variant, previously reported by Guda et al. [2009] to reduce enzymatic activity to 37% of the wild-type protein. One individual harbored a novel variant, c.1187A>G (p.Tyr396Cys).

The proband of family 20444 (Fig. 1), diagnosed with rectal cancer at age 35 years, carried the c.907G>A (p.Asp303Asn) variant. Her sister, who was diagnosed with CRC at age 61 years, carries the same *GALNT12* variant, as do six of the sister's eight children who were available for testing. In this sibship, adenomatous polyps were identified in 2/6 offspring that had undergone clinical screening—both of whom are carriers (Supp. Table S2). All variants have been submitted to the LOVD CRC database (www.lovd.nl/GALNT12).

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Michael O. Woods, Discipline of Genetics, Room 4333, Health Sciences Centre, 300 Prince Philip Drive, St. John's, NL, Canada A1B 3V6. E-mail: mwoods@mun.ca

Contract grant sponsors: National Cancer Institute of Canada (18223, 18226); Canadian Institutes of Health Research (CRT-43821); Genome Canada (Atlantic Medical Genetics and Genomics Initiative).

Table 1. Summary Data from the CRC Probands with Deleterious Germline GALNT12 Variants

Study #	Variant[a]	Proband age of diagnosis (CRC)	Location of CRC	MSI testing result[b]	IHC testing result (MLH1, MSH2, MSH6)	Bethesda criteria[c]	PolyPhen[d]	SIFT[e]
1117	c.907G>A p.Asp303Asn	72	Rectum	MSS	Intact	2, 4	PD	NT
20444	c.907G>A p.Asp303Asn	35	Rectum	ND	ND	4, 5	PD	NT
20896	c.907G>A p.Asp303Asn	72	Transverse Colon	MSS	Intact	2, 5	PD	NT
1728	c.1187A>G p.Tyr396Cys	21	Cecum	MSS	Intact	1	PD	NT

[a]Variant nomenclature is based upon the reference sequence NM_024642.4. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence.
[b]MSS, microsatellite stable; ND, not determined; but DNA sequencing of *MLH1*, *MSH2*, and *MSH6* did not reveal any deleterious variants;
[c]revised Bethesda criteria 1, CRC diagnosis at less than 50 years of age; 2, synchronous or metachronous tumors (colorectal or HNPCC-related tumor); 4, one first-degree relative with HNPCC-related tumor under 50 years of age; 5, two first- or second-degree relatives with HNPCC-related tumors at any age.
[d]PD, probably damaging.
[e]NT, not tolerated. Additional nonpathogenic variants were also identified (Supp. Table S1).

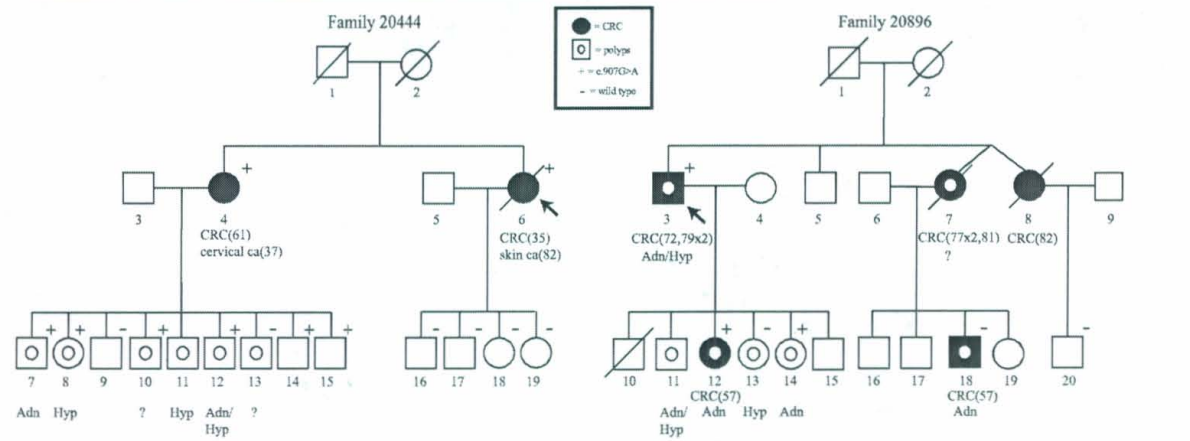

Figure 1. Two families segregating the c.907G>A GALNT12 variant. Arrows indicate probands. Numbers in brackets refer to ages of diagnoses (Hyp, hyperplastic polyp; Adn, adenomatous polyp; ?, polyp type unknown). Additional clinical information can be found in Supp. Table S2.

The proband of family 20896 (Fig. 1) also carries the c.907G>A variant and he presented with multiple polyps and was first diagnosed with CRC at age 72 years. Two daughters also carried the variant: one (PID12) developed CRC at age 57 and the second (PID14) has had two adenomas removed, both at age 45 years. A third daughter (PID13) did not carry the variant and had two hyperplastic polyps at 54 years of age. DNA was not available from two CRC-affected sisters of the proband, although a son of one of them (PID18) had CRC at age 57 years, yet did not carry the *GALNT12* variant and may thus represent a phenocopy.

Neither family 20444 nor 20896 met the Amsterdam criteria, possibly because screening colonoscopies identified precancerous polyps in several of those at risk.

The third proband (family 1117) harboring the c.907G>A (p.Asp303Asn) variant also fulfills the revised Bethesda criteria and had two CRCs at the age of 72 years, followed by skin cancer at the age of 74 years. The brother of the proband had CRC at age 47 years; however, no DNA was available for testing.

A novel variant, c.1187A>G (p.Tyr396Cys), was identified in the proband of a fourth family (family 1728). He also fulfills the revised Bethesda criteria and was diagnosed with CRC at 21 years of age. The father of the proband has not been diagnosed with cancer; however, the proband's paternal grandmother and his paternal great uncle both were diagnosed with CRC at age 70 years. Unfortunately, DNA was not available for any family member other than the proband, so segregation analysis was not possible. While no functional testing data is available, we conducted bioinformatic analyses on this variant and it was predicted to be pathogenic (PolyPhen [Sunyaev et al., 2001], SIFT [Ng and Henikoff, 2003]). Like p.Asp303Asn, p.Tyr396Cys lies in an important functional domain (catalytic) of the protein.

This is the first study to identify inherited deleterious *GALNT12* variants that segregate within families that appear to have an inherited form of CRC. We identified such variants in 4/118 (3.4%) families studied. We then sequenced *GALNT12* exons four and six in DNA from 149 healthy individuals from the NL population. The controls were initially contacted by phone through random digit dialing and were obtained for a population-based CRC study [Woods et al., 2010]. They are all Caucasian, from NL, consented to have blood drawn, completed a family history questionnaire and had no known history of any cancer in first-degree realtives. Neither variant was found in the controls, $P = 0.0376$ (Fisher's exact test). To determine if segregation of these variants in families 20444 and 20896 occurred by chance, we employed simplified rapid segregation analysis (SISA) [Moller et al., 2011]. Thus, we determined that the probability by chance that cosegregation of c.907G>A with CRC and/or adenomatous polyps occurred, in both pedigrees combined, was 1.56% (6.25% for 20444 and 25% for 20896).

Multiple linkage and association studies [Gray-McGuire et al., 2010; Kemp et al., 2006; Skoglund et al., 2006; Wiesner et al., 2003] have together provided strong evidence for a CRC-susceptibility gene in the chromosome region where *GALNT12* is located. This

study, together with the initial report from Guda et al. [2009], suggest that deleterious alleles in *GALNT12* can account for a proportion of that considerable group of CRC families in which the known CRC-associated genes have been excluded. Such variants may also increase the risk of other epithelial cancers [Guda et al., 2009].

It is possible that more extensive analyses, such as scanning for large exonic insertions/deletions or for regulatory variants will uncover additional families harboring deleterious variants in our clinical cohort. Future studies should focus on estimating the frequency of *GALNT12* variants in large population-based cohorts and should also estimate the penetrance in families that harbor deleterious variants in *GALNT12*.

Acknowledgments

We would like to extend our gratitude to the patients and their families who participated in this study. Without their interest and help none of this would be possible. Also, we thank Michelle Simms for her technical assistance and Dr. Bridget Fernandez for reviewing the manuscript.

References

Canadian Cancer Society's Steering Committee on Cancer Statistics. 2011. Canadian cancer statistics 2011. Toronto, ON: Canadian Cancer Society.

Gray-McGuire C, Guda K, Adrianto I, Lin CP, Natale L, Potter JD, Newcomb P, Poole EM, Ulrich CM, Lindor N, Goode EL, Fridley BL, Jenkins R, Le Marchand L, Casey G, Haile R, Hopper J, Jenkins M, Young J, Buchanan D, Gallinger S, Adams M, Lewis S, Willis J, Elston R, Markowitz SD, Wiesner GL. 2010. Confirmation of linkage to and localization of familial colon cancer risk haplotype on chromosome 9q22. Cancer Res 70:5409–5418.

Green RC, Green JS, Buehler SK, Robb JD, Daftary D, Gallinger S, McLaughlin JR, Parfrey PS, Younghusband HB. 2007. Very high incidence of familial colorectal cancer in Newfoundland: a comparison with Ontario and 13 other population-based studies. Fam Cancer 6:53–62.

Guda K, Moinova H, He J, Jamison O, Ravi L, Natale L, Lutterbaugh J, Lawrence E, Lewis S, Willson JK, Lowe JB, Wiesner GL, Parmigiani G, Barnholtz-Sloan J, Dawson DW, Velculescu VE, Kinzler KW, Papadopoulos N, Vogelstein B, Willis J, Gerken TA, Markowitz SD. 2009. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. Proc Natl Acad Sci USA 106:12921–12925.

Guo JM, Zhang Y, Cheng L, Iwasaki H, Wang H, Kubota T, Tachibana K, Narimatsu H. 2002. Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, pp-GalNAc-T12. FEBS Lett 524:211–218.

Kemp ZE, Carvajal-Carmona LG, Barclay E, Gorman M, Martin L, Wood W, Rowan A, Donohue C, Spain S, Jaeger E, Evans DG, Maher ER, Bishop T, Thomas H, Houlston R, Tomlinson I, Colorectal Tumour Gene Identification Study Consortium. 2006. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. Cancer Res 66:5003–5006.

Møller P, Clark N, Mæhle L. 2011. A SImplified method for Segregation Analysis (SISA) to determine penetrance and expression of a genetic variant in a family. Hum Mutat 32:568–571.

Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31:3812–3814.

Skoglund J, Djureinovic T, Zhou XL, Vandrovcova J, Renkonen E, Iselius L, Bisgaard ML, Peltomaki P, Lindblom A. 2006. Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1. J Med Genet 43:e7.

Sunyaev S, Ramensky V, Koch I, Lathe III W, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. Hum Mol Genet 10:591–597.

Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 96:261–268.

Vasen HF, Mecklin JP, Khan PM, Lynch HT. 1991. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 34:424–425.

Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC, Markowitz SD. 2003. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proc Natl Acad Sci USA 100:12961–12965.

Woods MO, Younghusband HB, Parfrey PS, Gallinger S, McLaughlin J, Dicks E, Stuckless S, Pollett A, Bapat B, Mrkonjic M, de la Chapelle A, Clendenning M, Thibodeau SN, Simms M, Dohey A, Williams P, Robb D, Searle C, Green JS, Green RC. 2010. The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. Gut 59:1369–1377.