HHS Public Access

Author manuscript

*Hum Mutat*. Author manuscript; available in PMC 2019 August 01.

Published in final edited form as:

*Hum Mutat*. 2018 August ; 39(8): 1092–1101. doi:10.1002/humu.23549.

# Evidence for *GALNT12* as a moderate penetrance gene for colorectal cancer

**Daniel R. Evans**[1,†], **Srividya Venkitachalam**[2,†], **Leslie Revoredo**[3], **Amanda T. Dohey**[1], **Erica Clarke**[1], **Julia J. Pennell**[1], **Amy E. Powell**[1], **Erina Quinn**[2], **Lakshmeswari Ravi**[2], **Thomas A. Gerken**[3,4,5], **Jane S. Green**[1], **Michael O. Woods**[1,*], and **Kishore Guda**[2,*]

[1]Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL A1B 3V6 Canada

[2]Division of General Medical Sciences-Oncology, Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland OH 44106, U.S.A

[3]Department of Chemistry, Case Western Reserve University, Cleveland OH 44106, U.S.A

[4]Department of Pediatrics, Case Western Reserve University School of Medicine, Cleveland OH 44106, U.S.A

[5]Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland OH 44106, U.S.A

## Abstract

Characterizing moderate penetrance susceptibility genes is an emerging frontier in colorectal cancer (CRC) research. *GALNT12* is a strong candidate CRC-susceptibility gene given previous linkage and association studies, and inactivating somatic and germline alleles in CRC patients. Previously, we found rare segregating germline *GALNT12* variants in a clinic-based cohort (*N*=118) with predisposition for CRC. Here, we screened a new population-based cohort of incident CRC cases (*N*=479) for rare (MAF  1%) deleterious germline *GALNT12* variants. *GALNT12* screening revealed 8 rare variants. Two variants were previously described (p.Asp303Asn, p.Arg297Trp), and additionally, we found 6 other rare variants: five missense (p.His101Gln, p.Ile142Thr, p.Glu239Gln, p.Thr286Met, p.Val290Phe) and one putative splice-altering variant (c.732-8 G>T). Sequencing of population-matched controls (*N*=400) revealed higher burden of variants in CRC cases compared to healthy controls (P=0.0381). We then functionally characterized the impact of these substitutions on GALNT12 enzyme activity using *in vitro*-derived peptide substrates. Three of the newly identified *GALNT12* missense variants (p.His101Gln, p.Ile142Thr, p.Val290Phe) demonstrated a marked loss (>2-fold reduction) of enzymatic activity compared to wild-type (P  0.05), while p.Glu239Gln exhibited a ~2-fold reduction in activity (P=0.077). These findings provide strong, independent evidence for the

Corresponding Author: Michael O. Woods, Associate Professor, Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, Craig L. Dobbin Genetics Research Centre, Rm 5M304, 300 Prince Phillip Drive, St. John's, NL, Canada, A1B 3V6, Phone: 709-864-6489, mwoods@mun.ca.

†Contributed equally to this work

*Contributed equally to this work

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

association of *GALNT12* defects with CRC-susceptibility; underscoring implications for glycosylation pathway defects in CRC.

**Keywords**

Colorectal cancer; *GALNT12*; Newfoundland & Labrador; Familial Colorectal Cancer Type X

## Introduction

Unraveling the genetic predisposition to colorectal cancer (CRC) is a subject of global significance (Rafiemanesh et al., 2016; Valle, 2014). Since the elucidation of mismatch repair (MMR) genes in hereditary CRC, the search for new and highly penetrant CRC-susceptibility genes has yielded diminishing returns. Indeed, recent studies now postulate that the existence of any major high penetrance susceptibility genes, of similar caliber as the MMR genes and *APC*, is unlikely (Chubb et al., 2016). This underscores the increasing relevance of CRC-susceptibility models that recognize the impact of rare alleles with modest effect sizes; which are difficult to detect by classic approaches such as genome-wide association and family-based linkage studies, yet could explain a sizeable fraction of the missing heritability in CRC (Hahn et al., 2016; Manolio et al., 2009). The genetic investigation of large and well-characterized patient cohorts is a useful way to identify moderate penetrance variants. While high penetrance variants are often the most clinically actionable, CRC screening guidelines now recognize moderate penetrance alleles in *APC*, *CHEK2* and *MUTYH* (Tung et al., 2016). With the genetic architecture of CRC becoming increasingly apparent, and the role of moderate penetrance genes becoming a new focus; this study explores *GALNT12* (MIM #610290, RefSeq: NM_024642.4) as a moderate penetrance gene for CRC susceptibility.

The initiation of mucin-type *O*-linked glycosylation is governed by the GALNT family of enzymes (Brockhausen, Schachter, & Stanley, 2009), and glycosylation plays a vital role in regulating key cellular functions including adhesion, migration and immune surveillance (Ohtsubo & Marth, 2006). PpGalNAc-T12 (encoded by *GALNT12*), is highly expressed in normal colon tissue and downregulated in colonic cancers (Guo et al., 2002; Guo, Chen, Wang, Zhang, & Narimatsu, 2004). Aberrant glycosylation can impair the normal functioning of lipids and proteins by affecting cellular glycan structure (Pinho & Reis, 2015; Stowell, Ju, & Cummings, 2015) and is a hallmark of several different cancers; including CRC (Bergstrom & Xia, 2013; Brockhausen, 2006). It is unsurprising, therefore, that genetic and epigenetic defects underlying the *O*-linked pathway are proposed to contribute to CRC pathogenesis (Brockhausen, 2006; Brockhausen et al., 2009; Tran & Ten Hagen, 2013).

Genetic evidence demonstrates that *GALNT12* is a candidate CRC-susceptibility gene; as we and others have identified a CRC-susceptibility locus spanning 9q22.32-31.1(Gray-McGuire et al., 2010; Skoglund et al., 2006; Wiesner et al., 2003) and *GALNT12* resides in the 6.5cM critical interval (markers D9S1851-D9S277) within this region (Kemp et al., 2006). Bolstered by functional evidence showing that inactivating *GALNT12* alleles are over-represented in CRC cell lines (Guda et al., 2009), we previously characterized

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

*GALNT12* in a clinic-based discovery cohort (*N*=118) with predilection for CRC, and identified rare and segregating *GALNT12* alleles in high-risk CRC families (Clarke et al., 2012).

Here, our aim was to expand the investigation of *GALNT12* by screening for rare and deleterious alleles in a larger and well-characterized population-based cohort of incident CRC cases (*N*=479) from the province of Newfoundland & Labrador (NL). This island of Atlantic Canada harbors a unique population of mostly English and Irish descent that is well-suited for genetic research; given its geographic and historical isolation (Mannion, 1977), and willingness to participate in research studies. With one of the highest reported incidence rates of familial CRC in the world (Green et al., 2007), and the highest incidence rate of CRC reported in Canada (Canadian Cancer Society's Advisory Committee on Cancer Statistics, 2015), the NL population offers an invaluable opportunity to study the genetic etiology of CRC (Rahman et al., 2003; Zhai et al., 2016). This was most significantly demonstrated by 'Family C', an NL kindred that helped elucidate the role of *MSH2* in hereditary CRC (Leach et al., 1993).

To further our investigation into the putative association of *GALNT12* defects and CRC susceptibility, by leveraging the unique attributes of the NL population, we examined the germline DNAs of 479 incident CRC cases from the Newfoundland Colorectal Cancer Registry (NFCCR) (Green et al., 2007). Herein, we describe novel findings of rare and deleterious *GALNT12* alleles in CRC patients, which cluster around the glycosyl-transferase domain, and functionally impair this enzyme.

# Methods

## Newfoundland Colorectal Cancer Registry and Population-matched Controls

During the prior NFCCR study, all pathologically confirmed incident CRC cases that occurred over a 5-year period (January 1st, 1999–December 31st, 2003), with diagnosis under the age of 75, were identified through provincial tumor registries (Green et al., 2007). After providing written and informed consent, study participants completed family history and risk factor questionnaires and provided access to blood samples, tumor tissues, medical records and gave permission for kin contact. As well, a cohort of CRC-free population controls was ascertained through random digit dialing (Raptis et al., 2007; Wang et al., 2009). After providing written and informed consent, population control participants provided family history, personal history and diet questionnaires, as well as blood samples for DNA extraction.

Patients of the NFCCR cohort have previously been screened for MMR gene variants and microsatellite instability (MSI) status (Woods et al., 2010). MMR gene variants were assessed for pathogenicity using the InSiGHT variant database (Thompson et al., 2014). We assigned familial CRC risk criteria based on relationship to the proband of each family. Risk criteria included in this study were the Amsterdam Criteria-1 (AC-1) (Vasen, Mecklin, Khan, & Lynch, 1991), the Age and Cancer Modified Amsterdam Criteria (ACMAC) (Woods et al., 2005), the Revised Bethesda Criteria (Umar et al., 2004), and criteria for Familial Colorectal Cancer Type X (FCCTX) (Lindor, 2009).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

During the current study, we chose germline DNAs of 479 NFCCR cases and 400 population-matched CRC-free controls, which then underwent *GALNT12* screening. We were blind to both MMR variant and microsatellite status during the study. The Health Research Ethics Authority of Newfoundland & Labrador granted approval for this study (HREB # 2010.035).

### Sanger DNA Sequencing and Variant Interpretation

CRC cases and population controls were sequenced using Sanger sequencing. Primer design was followed as per previous studies (Guda et al., 2009) and PCR amplification was performed for all 10 exons and intronic boundaries of *GALNT12* (conditions available upon request). The *GALNT12* RefSeq (NM_024642.4) and the canonical transcript (ENST00000375011.3) were selected using Ensembl genome browser, with the GRCh38.p10 reference genome assembly (Zerbino et al., 2018). Standard sequencing protocols for automated sequencing on an ABI 3130XL Genetic Analyzer (Applied Biosystems by Life Technologies, CA, USA) were followed. Sequencing data was then analyzed and interpreted using Sequencing Analysis 5.2 (Applied Biosystems by Life Technologies, CA, USA) and Sequencher 4.9 (Gene Codes Corporation, Michigan, USA).

Minor allele frequency (MAF) was assessed using publically available databases including: dbSNP (Sherry et al., 2001), the NHLBI exome variant server (NHLBI GO Exome Sequencing Project, Seattle, WA; URL http://evs.gs.washington.edu/EVS), the ExAC browser and the GnomAD browser (Lek et al., 2016). We calculated the maximum credible allele frequency of a CRC variant and variant penetrance using GnomAD frequencies (Whiffin et al., 2017) (Supp. Methods). A *GALNT12* variant MAF threshold of 1% in more than one population database was used to identify high impact variants of variable penetrance.

For each variant, bioinformatics tools including SIFT (Sim et al., 2012), Polyphen-2 (Adzhubei, Jordan, & Sunyaev, 2013), CADD (Kircher et al., 2014) and REVEL (Ioannidis et al., 2016) were used to predict functional consequences of rare variants. The putative splice-altering variant was assessed using Spliceman (http://fairbrother.biomed.brown.edu/spliceman/index.cgi), Human Splicing Finder (http://www.umd.be/HSF/), and SpliceView (http://bioinfo.itb.cnr.it/~webgene/wwwspliceview.html). Further details including program inputs and settings are provided in Supp. Methods. We submitted rare *GALNT12* variants to the Leiden Open Variation Database (LOVD) (www.lovd.nl/GALNT12).

### Generation of Secreted pIHV Constructs

A *GALNT12* cDNA fragment excluding the N-terminal transmembrane domain, corresponding to amino acids 38–581, was amplified by RT-PCR from RNA of a normal control as described previously (Guda et al., 2009). The resulting PCR fragments were cloned into *pIHV*, a modified SV40 promoter-driven pZeoSV2 vector (Life Technologies, Carlsbad, CA, USA) that contains an insulin secretion signal to direct the secretion of the recombinant protein into the cell culture medium and an N-terminal His6 and V5 epitope tags to facilitate purification and detection of the recombinant protein. cDNA fragments encoding mutant GALNT12 (p.His101Gln, p.Ile142Thr, p.Glu239Gln, p.Thr286Met,

Author Manuscript



p.Val290Phe, p.Arg297Trp, p.Asp303Asn and p.Tyr396Cys) were generated by site directed mutagenesis (QuikChange Lightning, Agilent Technologies, Santa Clara, CA, USA).

## Cell lines and DNA Transfection

SW480 cells were obtained from American Type Culture Collection (ATCC) in 2012, tested for authenticity using short tandem repeat genotyping at least 3 months prior to experimental assessments, and screened periodically for mycoplasma contamination using the MycoAlert mycoplasma detection kit (Lonza, Basel, Switzerland). Transfection was performed using Lipofectamine 2000 (Life Technologies, Carlsbad, CA, USA) according to standard protocol. Briefly, $10^6$ SW480 cells were plated per 100-mm dish 24 hours before transfection, and incubated in 5% $CO_2$ at 37 °C overnight. 4 μg of plasmid DNA was used per 100-mm dish.

## Recombinant Protein Purification

The cell monolayers were washed twice with ice-cold PBS and incubated with lysis buffer (50mMTris, pH 7.5/150mMNaCl/1mMCaCl2/1mM MnCl2/EDTA-free protease inhibitor pellets/ 0.3% CHAPS) for 15 min on ice. After scraping, the lysates were clarified by centrifugation for 15 min at maximal speed. The recombinant protein was immunoprecipitated from the lysates using anti-V5 agarose beads (Sigma-Aldrich, St Louis, MO, USA), and subsequently washed with wash buffer (50mMTris, pH 7.5/150mMNaCl/1mMCaCl2/1mM MnCl2/EDTA-free protease inhibitor pellets).

## Western Blot Analysis

After immunoprecipitation, 1/10 fraction of the recombinant protein was mixed with equal volume of Laemmli sample buffer (Bio-Rad, Hercules, CA, USA) at 95 °C for 5 min, and loaded onto a Bis-Tris SDS/4–12% polyacrylamide gel (Life Technologies, Carlsbad, CA, USA). After SDS/PAGE, proteins were transferred onto Immobilon-P PVDF membranes (EMD Millipore, Billerica, MA). Membranes were blocked for 1 hr with 5% nonfat milk, and incubated with appropriate dilution of mouse anti-V5 antibody conjugated to horseradish peroxidase (Life Technologies, Carlsbad, CA, USA) to detect the V5-tagged proteins. Enhanced Chemiluminescence Plus (GE Healthcare-BioSciences, Pittsburg, PA, USA) and ImageJ software (National Institutes of Health, Bethesda, MA, USA) were used to detect and quantitate respective protein bands.

## Enzymatic Assay for *GALNT12* Variants

Negative control (empty vector), positive control (wild-type GALNT12), and missense mutant versions of GALNT12-bound beads were added to transferase-specific reaction mixtures in 1.5mL Eppendorf microcentrifuge tubes. Transferase-specific reaction mixtures contained 40 mM sodium cacodylate pH 6.8, 0.32 mM 2-mercaptoethanol, 0.03% Triton-X 100, 10 mM $MnCl_2$, 1 mM UDP-GalNAc containing 0.1 *u*Ci of [$^3$H] UDP-GalNAc (American Radiolabeled Chemicals Inc., St. Louis, MO, USA), protease inhibitors A and B (Sigma-Aldrich, St Louis, MO, USA) and 1 mM (0.5 mg/mL) OPT-T12 substrate GAGAYYITPRPGAGA (RS Synthesis, Louisville, KY). Of note, OPT-T12 is an optimal peptide substrate specifically determined for ppGalNAc T12 by the Gerken Lab (Gerken et

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

al., 2011) and custom synthesized by RS Synthesis. A 10 mmol stock solution of OPT-T12 substrate was prepared by lyophilizing from water several times and adjusting to pH 7.4 with dilute NaOH/HCL. Reagent reaction mixtures were combined with 100 μL transferase-bound beads to a final reaction volume of 250 μL. The pH of the total reaction was adjusted to 6.8 and the mixture was agitated at 37°C in a TAITEC shaking Microincubator M-36 to maintain the beads in suspension. Following an overnight incubation, reaction mixtures were quenched with an equal volume of 250 mM EDTA and frozen for later processing. UDP and non-hydrolyzed UDP-GalNAc were removed by passing the sample through a column of ~3mL of Dowex 1x8 anion exchange resin (Acros Organics, Thermo Fisher Scientific, Pittsburg, PA). Subsequently, eluents were lyophilized, reconstituted with water to a final volume of 1ml and passed through G10 columns. [$^{3}$H]-GalNAc incorporation onto the OPT-T12 substrate was determined by scintillation counting on a Beckman LS5801 scintillation counter while absorbance at 220 nm and 280 nm was measured using a spectrophotometer. Transferase activity at each time point was calculated as a ratio of the post-G10 counts to the absorbance at 220 nm. Mutant transferase-specific activities were then normalized to the WT transferase to yield an approximate relative specific activity of each mutant. The specific activities of the wild-type, and each of the GALNT12 mutants, were assayed and were expressed as the average activity derived from 3 independent biological replicates per experimental arm.

### Statistical Analyses

Transferase activities of wild-type and mutant GALNT12 proteins were compared using a Student’s t-test, with a P value 0.05 considered statistically significant. A one-sided Fisher’s exact test was used to compare the frequency of *GALNT12* variants in CRC cases and population-matched controls, with a P value of 0.05 considered statistically significant.

## Results

Germline DNA sequencing of 479 NFCCR cases revealed 21 polymorphisms in *GALNT12* (Supp. Table S1). Analysis of rare (MAF 1%) *GALNT12* variants identified 8 candidates for further study (Table 1). These included 7 missense variants (c.303C>G, p.His101Gln; c.425T>C, p. Ile142Thr; c.715G>C, p.Glu239Gln; c.857C>T, p.Thr286Met; c.868G>T, p.Val290Phe; c.889C>T, p.Arg297Trp; c.907G>A, p.Asp303Asn) and 1 putative splice-altering variant (c.732-8G>T) (Figure 1). GnomAD MAFs ranged from 0.1244% (p.Asp303Asn) (344 alleles) to 0.0004061% (p.Glu239Gln) (1 allele). Maximum credible allele frequency was calculated under different CRC prevalence and penetrance assumptions to explore potential frequency thresholds (Supp. Table S2). The 8 candidate *GALNT12* variants were sequenced in a population-matched CRC-free control cohort (*N*=400). Control sequencing revealed these 8 variants were over-represented in CRC cases compared to population-matched CRC-free controls (P = 0.03814), with the p.Asp303Asn and p.Arg297Trp variants each found in a single control respectively (Table 1, Supp. Figure S1). Multiple *in silico* tools (SIFT, Polyphen, CADD, REVEL) were used to predict consequences of the *GALNT12* variants (Table 1). Missense variants scored deleterious according to CADD, SIFT and Polyphen-2. CADD scores were highly deleterious, ranging from 24.8 (p.Asp303Asn) to 34.0 (p.Thr286Met).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



*GALNT12* variants clustered within or near the glycosyltransferase domain encoded by *GALNT12* (Figure 2). Transferase activity of missense variants was characterized using an *in vitro*-derived peptide substrate assay (Figure 3). Three of the newly identified *GALNT12* missense variants (p.His101Gln, p.Ile142Thr, p.Val290Phe) demonstrated a marked loss (>2-fold reduction) of enzymatic activity compared to the wild-type protein (P  0.05), while the p.Glu239Gln variant exhibited a ~2-fold reduction in enzymatic activity (P=0.077). Additionally, we tested the p.Tyr396Cys allele identified in our previous study (Clarke et al., 2012), which demonstrated a significant loss in GALNT12 activity (P  0.05).

*In silico* tools (Spliceman, Human Splicing Finder, Spliceview) were used to predict consequences of the putative splice-altering variant (c.732-8G>T), since RNA was not available because patients live in a rural and remote area (Supp. Methods). Predictions were discordant, as Spliceman predicted high probability of disrupted splicing (L1 81%); whereas Human Splicing Finder predicted alteration of an exonic splice silencing site or alternatively no effect; while Spliceview predicted no influence on splicing.

The 8 rare *GALNT12* variants were derived from probands of 10 unrelated families, with p.Asp303Asn represented in 3 unrelated families (Figure 1). We proceeded with familial co-segregation analysis to assess penetrance. DNA was not available for segregation analysis in 9 out of 10 of families, and many relatives were deceased at the time of study. Segregation was tested in Family 1117 (Figure 1), as DNA was available for two distant CRC-affected relatives of the proband (first cousins once removed). Neither carried the p.Asp303Asn variant, and it is unclear whether the proband's risk for CRC derives from the paternal or maternal lineages. Our limited co-segregation analysis was inconclusive, but not supportive of p.Asp303Asn segregation (Supp. Figure S2). Given limitations of DNA available for segregation analysis, we estimated penetrance (Minikel et al., 2016; Whiffin et al., 2017) under different assumptions specific to the NL population (Supp. Methods), observing evidence of moderate penetrance for *GALNT12* variants (Supp. Table S2).

Table 2 shows the clinical and molecular characteristics of CRC probands carrying rare *GALNT12* variants. The age of CRC diagnosis ranged from 43–72 years. Patients of the NFCCR cohort were screened for MMR defects or MSI during prior studies (Woods et al., 2010), and we had no *a priori* knowledge of this status during this study. We observed that the majority of probands with rare *GALNT12* variants (8 out of 10) were microsatellite stable (MSS) and MMR proficient. Probands of two families (Figure 1: 1575, 1595) harbored pathogenic (InSiGHT class 5) MMR gene variants with high tumor microsatellite instability (MSI-H) (Table 2). Finally, we assessed risk of genetic predisposition for CRC in all families using established risk criteria. Two families (Figure 1: 1595, 1541) fulfilled the AC-1 (Vasen et al., 1991). In the two AC-1 families, the proband of Family 1595 was MSI-H with MLH1 deficiency and the proband of Family 1541 was MSS and MMR proficient, the latter thereby meeting criteria for FCCTX. A total of five families met the Revised Bethesda Criteria (Umar et al., 2004). Two were the AC-1 families described above, and of the remaining three, one was MSI-H *MSH6* deficient, and the remaining two were MSS and MMR intact.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript



## Discussion

Here, we screened a population-based cohort of 479 incident CRC cases, identifying 8 rare *GALNT12* variants (Table 1). We observed a higher burden of these variants in CRC cases compared to population-matched CRC-free controls (P   0.05). Multiple *in silico* tools (CADD, SIFT and Polyphen-2) predicted these missense variants are deleterious (Table 1), and variants clustered within or near the functionally relevant glycosyltransferase domain (Figure 2). Transferase assays revealed these GALNT12 mutants demonstrate a marked reduction of enzyme activity compared to wild-type (Figure 3).

To first explore *GALNT12* variation, we chose a rare variant MAF threshold of   1%. This MAF allowed us to assess variants under reduced penetrance, while including the potential for founder effects in the NL population. Recently, Whiffin et al proposed a strategy that leverages allele frequencies from large population databases to assess maximal credible allele frequencies (Whiffin et al., 2017). Predicted maximum credible allele frequencies for a CRC variant in NL are shown under different models of penetrance in Supp. Table S2. For example, under 50% penetrance, we estimated a MAF of no greater than 0.00997%, or 37 GnomAD alleles (Supp. Methods). During this analysis, we noted that the most common known pathogenic CRC variant in NL, a founder variant in *MSH2* (p.Val265_Gln314del), greatly exceeds our calculated maximum credible allele frequencies, with a MAF of 1.125% (330 alleles out of 29332 total alleles; 18 homozygotes) in GnomAD. This pathogenic variant is known to cause CRC in multiple families around the world and explains approximately 37% of solved CRC cases in NL (Woods et al., 2010). This finding likely arises from a very high number of false positive calls due to low sequence complexity in this region; highlighting a confounding variable in using these large databases, which rely on next-generation DNA sequencing. We would therefore advocate for a judicious and cautious approach to relying on calculated frequencies from these databases as a sole indicator of pathogenicity.

Bioinformatic tools (SIFT, Polyphen-2) are well-established for predicting consequences of genetic variation (Dong et al., 2015). SIFT assumes variation within evolutionarily conserved regions is more likely to impact protein function (Sim et al., 2012), while Polyphen-2 considers sequence-based, phylogenetic and structural features to predict consequences (Adzhubei et al., 2013). We observed agreement between SIFT and Polyphen-2 in predicting deleterious consequences for the *GALNT12* missense variants (Table 1). The sole exception was p.Asp303Asn; which SIFT predicted ‘tolerated’ while Polyphen-2 predicted ‘probably damaging’. A limitation of these tools is that they cannot predict intronic variation. Newer ensemble prediction tools such as CADD and REVEL are gaining favor since they incorporate weighted estimates using multiple *in silico* tools, trained on machine-learning algorithms, to assist in variant interpretation (Ioannidis et al., 2016; Kircher et al., 2014). CADD scores have demonstrated clinical utility for interpreting MMR variants in a manually-curated dataset from InSiGHT (van der Velde et al., 2015). We observed highly deleterious CADD scores (24.8–34.0) for all *GALNT12* missense variants (Table 1). REVEL scores predicted two of the *GALNT12* variants (p.Ile142Thr, p.Glu239Gln) are deleterious (>0.5); thus CADD and REVEL were not concordant. Though both tools have merits, they are limited by circular errors arising from the influence of their

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

respective benchmarking datasets (Grimm et al., 2015). Overall, the rare missense *GALNT12* variants were predicted deleterious according to three programs -- CADD, SIFT and Polyphen-2. Since these deleterious variants clustered around the glycosyltransferase domain (Figure 2), we postulated that assaying enzymatic activity might further delineate the consequences of these variants.

Accordingly, we characterized mutant enzyme activities using an *in-vitro* peptide derived assay (Figure 3). We observed that GALNT12 mutants showed a  2-fold reduction of enzyme activity in 7 of the variants (including the previously identified p.Tyr396Cys), with five mutants (p.His101Gln, p.Ile142Thr, p.Val290Phe, p.Arg297Trp, p.Tyr396Cys) reaching the threshold for statistical significance (P  0.05). GALNT12 expression was consistently higher in mutants compared to wild-type, which demonstrated the lowest protein expression (Figure 3). This observation suggests our assay might underestimate the true reduction of enzyme activity in the mutants. The consistent differences in GALNT12 expression could be the result of a compensatory mechanism arising from altered protein stability, and is an avenue for future studies. Taken together, our enzymatic assays provide corroborating evidence for the deleterious effects of these *GALNT12* variants.

We next pursed familial co-segregation analysis of the rare *GALNT12* variants to assess penetrance. Due to our ascertainment method (incident cases), additional DNA were not available for 9 out of 10 families; thus segregation could not be tested in most families. We tested segregation of p.Asp303Asn in Family 1117 (Figure 1). The results did not support segregation, though pedigree structure and limited DNA sample number rendered the analysis inconclusive (Supp. Figure S2). Recently, Whiffin et al. proposed a strategy to estimate variant penetrance using allele frequencies in large population databases (Minikel et al., 2016; Whiffin et al., 2017). Using the projected prevalence of CRC in NL (Supp. Methods), we estimated penetrance of *GALNT12* variants in Supp. Table S2, and found *in silico* evidence for moderate to low penetrance of these variants. Given that CRC is a late onset disorder, we also calculated the penetrance using the Canadian lifetime risk for developing CRC (1 in 13), and found stronger evidence for moderate penetrance. While these *in silico* estimates can provide a broad estimate of penetrance (orders of magnitude), they remain confounded by population stratification and sampling variance, especially for the very low frequency alleles in the population databases. A formal assessment of penetrance for *GALNT12* variants remains an important avenue for future studies, and argues the importance of pursuing these *GALNT12* variants in CRC families derived from diverse populations.

The tumor characteristics and molecular profiles of probands harboring *GALNT12* variants are shown in Table 2. Without *a priori* knowledge of MMR or MSI status, we observed most probands with a rare *GALNT12* variant (8 out of 10) were MSS with MMR stable. Two probands (Umar et al., 2004) (Figure 1: 1575, 1595) concomitantly carried one *GALNT12* variant (p.His101Gln or p.Thr286Met) with a pathogenic (InSiGHT class 5) MMR variant (*MSH6;* NM_000179.2: p.Arg1172Lysfs*5 or *MLH1* NM_000249.3: p.His264Leufs*2) respectively. GALNT12 p.His101Gln mutants demonstrated significant reduction in enzyme activity while p.Thr286Met mutants did not (Figure 3). It is difficult to assess the role of the *GALNT12* variants, as the concomitant pathogenic MMR variants undoubtedly contributed

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to CRC predisposition in these Lynch Syndrome families. One advantage to our population-based approach is that we did not exclude patients with prior known causes of CRC; which is an exclusionary bias often represented by studies exploring novel genes. It would be speculative to assume that the *GALNT12* variants are non-causal based on the concomitant MMR variants alone, and this reasoning overlooks the influence of population genetics. For example, there are previously documented cases of bilineal inheritance and trans-heterozygotes for autosomal dominant polycystic kidney disease segregating in the NL population (Pei et al., 2001). Thus, two pathogenic variants for a monogenic disease can, and do co-occur, within single families. Further study encompassing a broader array of populations will be useful to delineate the role of *GALNT12* variants.

In total, three non-Lynch Syndrome families met the Revised Bethesda Criteria, with one family meeting criteria for FCCTX (Table 2). Recently, germline DNA sequencing of *GALNT12* in a Spanish cohort of 103 FCCTX families failed to identify any rare candidate variants in this gene (Seguí et al., 2014). While only 45.6% of families in the Seguí *et al.* (2014) study met AC-1 and were truly FCCTX, the differences between this study and ours underscore the importance in recognizing the influence of population genetics as well as ascertainment criteria. A negative screening study from a single population should not pre-empt the conclusion of a lack of a role for *GALNT12* in CRC.

Strong arguments exist for the putative role of *GALNT12* in CRC pathogenesis. Genetic evidence demonstrates that *GALNT12* resides within a CRC susceptibility locus, with both rare and functionally deleterious somatic and germline *GALNT12* variants found in CRC patients (Clarke et al., 2012; Guda et al., 2009). Moreover, GALNT12 is selectively expressed in intestinal epithelia and frequently downregulated in CRCs (Guo et al., 2002, Guo et al., 2004), and has been identified as one of the key genes facilitating normal morphology and polarization in colon epithelial cells (Li et al., 2017). Additionally, GALNT12 is a key enzyme catalyzing the initiation of mucin type *O*-glycosylation (Bennett et al., 2012); a pathway shown to play a critical role in CRC progression (Pinho & Reis, 2015; Stowell et al., 2015). Mice deficient in intestinal mucin-type *O*-glycans spontaneously develop colitis-associated colon cancer due to inflammation resulting from impaired mucin barrier function (Bergstrom et al., 2016). Moreover, mice lacking Muc2 spontaneously develop adenomas and CRC (Velcich et al., 2002). Finally, recent studies of human glycosylation-associated genes from CRC-cell lines have identified candidates that are significantly mutated in CRC (Venkitachalam et al., 2016).

Taken together, our findings suggest a thorough investigation for any putative link between *GALNT12* and CRC progression is necessary. A broader approach for testing rare *GALNT12* alleles in diverse populations and characterizing their biological function in pre-clinical animal models may help decipher a tumorigenic role of *GALNT12* in CRC pathogenesis. The discovery of moderate penetrance genes for CRC is an important avenue to expand our understanding of the genetic predisposition for CRC. Although challenging, the characterization of these genes is poised to impart novel insights in the interpretation of cancer risk moving forward. The investigation of large and well-characterized cohorts from well-studied populations represents an important strategy moving forward.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Funding**

This work was funded by the Research & Development Corporation of Newfoundland & Labrador (RDCNL), the Canadian Cancer Society Research Institute (CCSRI), K08 CA148980 (K.G.); Case GI SPORE P50 CA150964 (K.G.); P30 CA043703 (K.G.); U54 CA163060 (K.G.); R01 CA204549 (K.G.); U01 CA152756 (K.G.); R01 CA078834 (T.A.G.) and U01 GM113534 (T.A.G.). The authors declare they hold no conflict of interest with respect to this work.

DRE and SV contributed equally to this manuscript as co-first authors. DRE prepared the manuscript and screened CRC patient DNA samples for *GALNT12* variants. SV performed the cloning, mutagenesis and recombinant protein purification with assistance from LaR and EQ. SV and LeR performed functional assays and SV assisted in drafting the manuscript. ATD, ECC, AEP and JJP participated in sequencing of patient DNA samples as past members of the MOW laboratory. TAG supervised all enzyme assays. KG and MOW supervised the collection of molecular and genetic data, contributed to study design and assisted in preparation and review of the manuscript. All authors read and approved the final manuscript.

We would like to acknowledge the late Dr. Roger C. Green, who worked extensively to establish and maintain the Newfoundland Colorectal Cancer Registry. We would like to thank the study participants who, after providing informed consent, gave their family histories and DNA samples for our current and ongoing studies.

## References

Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics. 2013; Chapter 7(Unit7.20)doi: 10.1002/0471142905.hg0720s76

Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology. 2012; 22(6):736–756. DOI: 10.1093/glycob/cwr182 [PubMed: 22183981]

Bergstrom K, Liu X, Zhao Y, Gao N, Wu Q, Song K, … Xia L. Defective Intestinal Mucin-Type O-Glycosylation Causes Spontaneous Colitis-Associated Cancer in Mice. Gastroenterology. 2016; 151(1):152–164e11. DOI: 10.1053/j.gastro.2016.03.039 [PubMed: 27059389]

Bergstrom KSB, Xia L. Mucin-type O-glycans and their roles in intestinal homeostasis. Glycobiology. 2013; 23(9):1026–1037. DOI: 10.1093/glycob/cwt045 [PubMed: 23752712]

Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. EMBO Reports. 2006; 7(6):599–604. DOI: 10.1038/sj.embor.7400705 [PubMed: 16741504]

Brockhausen I, , Schachter H, , Stanley P. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, , … Etzler ME, editorsEssentials of Glycobiology 2. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009 Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK1896/

Canadian Cancer Society's Advisory Committee on Cancer StatisticsCanadian Cancer Statistics 2015 Toronto, ON: Canadian Cancer Society; 2015

Chubb D, Broderick P, Dobbins SE, Frampton M, Kinnersley B, Penegar S, … Houlston RS. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nature Communications. 2016; 7:11883.doi: 10.1038/ncomms11883

Clarke E, Green RC, Green JS, Mahoney K, Parfrey PS, Younghusband HB, Woods MO. Inherited deleterious variants in GALNT12 are associated with CRC susceptibility. Human Mutation. 2012; 33(7):1056–1058. DOI: 10.1002/humu.22088 [PubMed: 22461326]

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Human Molecular Genetics. 2015; 24(8):2125–2137. DOI: 10.1093/hmg/ddu733 [PubMed: 25552646]

Gerken TA, Jamison O, Perrine CL, Collette JC, Moinova H, Ravi L, … Tabak LA. Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. The Journal of Biological Chemistry. 2011; 286(16): 14493–14507. DOI: 10.1074/jbc.M111.218701 [PubMed: 21349845]

Gray-McGuire C, Guda K, Adrianto I, Lin CP, Natale L, Potter JD, … Wiesner GL. Confirmation of linkage to and localization of familial colon cancer risk haplotype on chromosome 9q22. Cancer Research. 2010; 70(13):5409–5418. DOI: 10.1158/0008-5472.CAN-10-0188 [PubMed: 20551049]

Green RC, Green JS, Buehler SK, Robb JD, Daftary D, Gallinger S, … Younghusband HB. Very high incidence of familial colorectal cancer in Newfoundland: a comparison with Ontario and 13 other population-based studies. Familial Cancer. 2007; 6(1):53–62. DOI: 10.1007/s10689-006-9104-x [PubMed: 17039269]

Grimm DG, Azencott CA, Aicheler F, Gieraths U, MacArthur DG, Samocha KE, … Borgwardt KM. The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. Human Mutation. 2015; 36(5):513–523. DOI: 10.1002/humu.22768 [PubMed: 25684150]

Guda K, Moinova H, He J, Jamison O, Ravi L, Natale L, … Markowitz SD. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(31): 12921–12925. DOI: 10.1073/pnas.0901454106 [PubMed: 19617566]

Guo JM, Chen HL, Wang GM, Zhang YK, Narimatsu H. Expression of UDP-GalNAc:Polypeptide N-Acetylgalactosaminyltransferase-12 in Gastric and Colonic Cancer Cell Lines and in Human Colorectal Cancer. Oncology. 2004; 67(3–4):271–276. DOI: 10.1159/000081328 [PubMed: 15557789]

Guo JM, Zhang Y, Cheng L, Iwasaki H, Wang H, Kubota T, … Narimatsu H. Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family, pp-GalNAc-T12. FEBS Letters. 2002; 524(1–3):211–218. [PubMed: 12135769]

Hahn MM, de Voer RM, Hoogerbrugge N, Ligtenberg MJL, Kuiper RP, van Kessel AG. The genetic heterogeneity of colorectal cancer predisposition - guidelines for gene discovery. Cellular Oncology (Dordrecht). 2016; 39(6):491–510. DOI: 10.1007/s13402-016-0284-6 [PubMed: 27279102]

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, … Sieh W. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. American Journal of Human Genetics. 2016; 99(4):877–885. DOI: 10.1016/j.ajhg.2016.08.016 [PubMed: 27666373]

Kemp ZE, Carvajal-Carmona LG, Barclay E, Gorman M, Martin L, Wood W. … Colorectal Tumour Gene Identification Study Consortium. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. Cancer Research. 2006; 66(10):5003–5006. DOI: 10.1158/0008-5472.CAN-05-4074 [PubMed: 16707420]

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics. 2014; 46(3):310–315. DOI: 10.1038/ng.2892 [PubMed: 24487276]

Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, … Nyström-Lahti M. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell. 1993; 75(6):1215–1225. [PubMed: 8261515]

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T. … Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536(7616):285–291. DOI: 10.1038/nature19057 [PubMed: 27535533]

Li C, Singh B, Graves-Deal R, Ma H, Starchenko A, Fry WH, … Coffey RJ. Three-dimensional culture system identifies a new mode of cetuximab resistance and disease-relevant genes in colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America. 2017; 114(14):E2852–E2861. DOI: 10.1073/pnas.1618297114 [PubMed: 28320945]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Lindor NM. Familial colorectal cancer type X: the other half of hereditary nonpolyposis colon cancer syndrome. Surgical Oncology Clinics of North America. 2009; 18(4):637–645. DOI: 10.1016/j.soc.2009.07.003 [PubMed: 19793571]

Mannion JJ. The peopling of Newfoundland: essays in historical geography St. Johns, NL: Institute of Social and Economic Research, Memorial University of Newfoundland; 1977 Retrieved from http://collections.mun.ca/cdm/ref/collection/cns/id/34124

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, … Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. DOI: 10.1038/nature08494 [PubMed: 19812666]

Minikel EV, Vallabh SM, Lek M, Estrada K, Samocha KE, Sathirapongsasuti JF, … MacArthur DG. Quantifying prion disease penetrance using large population control cohorts. Science Translational Medicine. 2016; 8(322):322ra9.doi: 10.1126/scitranslmed.aad5169

Ohtsubo K, Marth JD. Glycosylation in cellular mechanisms of health and disease. Cell. 2006; 126(5):855–867. DOI: 10.1016/j.cell.2006.08.019 [PubMed: 16959566]

Pei Y, Paterson AD, Wang KR, He N, Hefferton D, Watnick T, … St George-Hyslop P. Bilineal disease and trans-heterozygotes in autosomal dominant polycystic kidney disease. American Journal of Human Genetics. 2001; 68(2):355–363. [PubMed: 11156533]

Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nature Reviews. Cancer. 2015; 15(9):540–555. DOI: 10.1038/nrc3982 [PubMed: 26289314]

Rafiemanesh H, Mohammadian-Hafshejani A, Ghoncheh M, Sepehri Z, Shamlou R, Salehiniya H, … Makhsosi BR. Incidence and Mortality of Colorectal Cancer and Relationships with the Human Development Index across the World. Asian Pacific Journal of Cancer Prevention: APJCP. 2016; 17(5):2465–2473. [PubMed: 27268615]

Rahman P, Jones A, Curtis J, Bartlett S, Peddle L, Fernandez BA, Freimer NB. The Newfoundland population: a unique resource for genetic investigation of complex diseases. Human Molecular Genetics, 12 Spec No. 2003; 2:R167–172. DOI: 10.1093/hmg/ddg257

Raptis S, Mrkonjic M, Green RC, Pethe VV, Monga N, Chan YM, … Bapat B. MLH1-93G>A promoter polymorphism and the risk of microsatellite-unstable colorectal cancer. Journal of the National Cancer Institute. 2007; 99(6):463–474. DOI: 10.1093/jnci/djk095 [PubMed: 17374836]

Seguí N, Pineda M, Navarro M, Lázaro C, Brunet J, Infante M, … Valle L. GALNT12 is not a major contributor of familial colorectal cancer type X. Human Mutation. 2014; 35(1):50–52. DOI: 10.1002/humu.22454 [PubMed: 24115450]

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research. 2001; 29(1):308–311. [PubMed: 11125122]

Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Research. 2012; 40(Web Server issue):W452–457. DOI: 10.1093/nar/gks539 [PubMed: 22689647]

Skoglund J, Djureinovic T, Zhou XL, Vandrovcova J, Renkonen E, Iselius L, … Lindblom A. 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. 2006; 43(2):e7.doi: 10.1136/jmg.2005.033928 [PubMed: 16467217]

Stowell SR, Ju T, Cummings RD. Protein glycosylation in cancer. Annual Review of Pathology. 2015; 10:473–510. DOI: 10.1146/annurev-pathol-012414-040438

Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F. … On Behalf Of InSiGHT. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nature Genetics. 2014; 46(2):107–115. DOI: 10.1038/ng.2854 [PubMed: 24362816]

Tran DT, Ten Hagen KG. Mucin-type O-glycosylation during development. The Journal of Biological Chemistry. 2013; 288(10):6921–6929. DOI: 10.1074/jbc.R112.418558 [PubMed: 23329828]

Tung N, Domchek SM, Stadler Z, Nathanson KL, Couch F, Garber JE, … Robson ME. Counselling framework for moderate-penetrance cancer-susceptibility mutations. Nature Reviews. Clinical Oncology. 2016; 13(9):581–588. DOI: 10.1038/nrclinonc.2016.90

Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, … Srivastava S. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

microsatellite instability. Journal of the National Cancer Institute. 2004; 96(4):261–268. [PubMed: 14970275]

Valle L. Genetic predisposition to colorectal cancer: where we stand and future perspectives. World Journal of Gastroenterology. 2014; 20(29):9828–9849. DOI: 10.3748/wjg.v20.i29.9828 [PubMed: 25110415]

van der Velde KJ, Kuiper J, Thompson BA, Plazzer JP, van Valkenhoef G, de Haan M. … InSiGHT Group. Evaluation of CADD Scores in Curated Mismatch Repair Gene Variants Yields a Model for Clinical Validation and Prioritization. Human Mutation. 2015; 36(7):712–719. DOI: 10.1002/humu.22798 [PubMed: 25871441]

Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Diseases of the Colon and Rectum. 1991; 34(5): 424–425. [PubMed: 2022152]

Velcich A, Yang W, Heyer J, Fragale A, Nicholas C, Viani S, … Augenlicht L. Colorectal cancer in mice genetically deficient in the mucin Muc2. Science (New York, NY). 2002; 295(5560):1726–1729. DOI: 10.1126/science.1069094

Venkitachalam S, Revoredo L, Varadan V, Fecteau RE, Ravi L, Lutterbaugh J, … Guda K. Biochemical and functional characterization of glycosylation-associated mutational landscapes in colon cancer. Scientific Reports. 2016; 6:23642.doi: 10.1038/srep23642 [PubMed: 27004849]

Wang PP, Dicks E, Gong X, Buehler S, Zhao J, Squires J, … Parfrey PS. Validity of random-digit-dialing in recruiting controls in a case-control study. American Journal of Health Behavior. 2009; 33(5):513–520. [PubMed: 19296741]

Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, … Ware JS. Using high-resolution variant frequencies to empower clinical genome interpretation. Genetics in Medicine: Official Journal of the American College of Medical Genetics. 2017; 19(10):1151–1158. DOI: 10.1038/gim.2017.26 [PubMed: 28518168]

Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, … Markowitz SD. A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(22):12961–12965. DOI: 10.1073/pnas.2132286100 [PubMed: 14566058]

Woods MO, Younghusband HB, Parfrey PS, Gallinger S, McLaughlin J, Dicks E, … Green RC. The genetic basis of colorectal cancer in a population-based incident cohort with a high rate of familial disease. Gut. 2010; 59(10):1369–1377. DOI: 10.1136/gut.2010.208462 [PubMed: 20682701]

Woods MO, Hyde AJ, Curtis FK, Stuckless S, Green JS, Pollett AF, … Parfrey PS. High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research. 2005; 11(19 Pt 1):6853–6861. DOI: 10.1158/1078-0432.CCR-05-0726 [PubMed: 16203774]

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, … Flicek P. Ensembl 2018. Nucleic Acids Research. 2018; 46(D1):D754–D761. DOI: 10.1093/nar/gkx1098 [PubMed: 29155950]

Zhai G, Zhou J, Woods MO, Green JS, Parfrey P, Rahman P, Green RC. Genetic structure of the Newfoundland and Labrador population: founder effects modulate variability. European Journal of Human Genetics: EJHG. 2016; 24(7):1063–1070. DOI: 10.1038/ejhg.2015.256 [PubMed: 26669659]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript



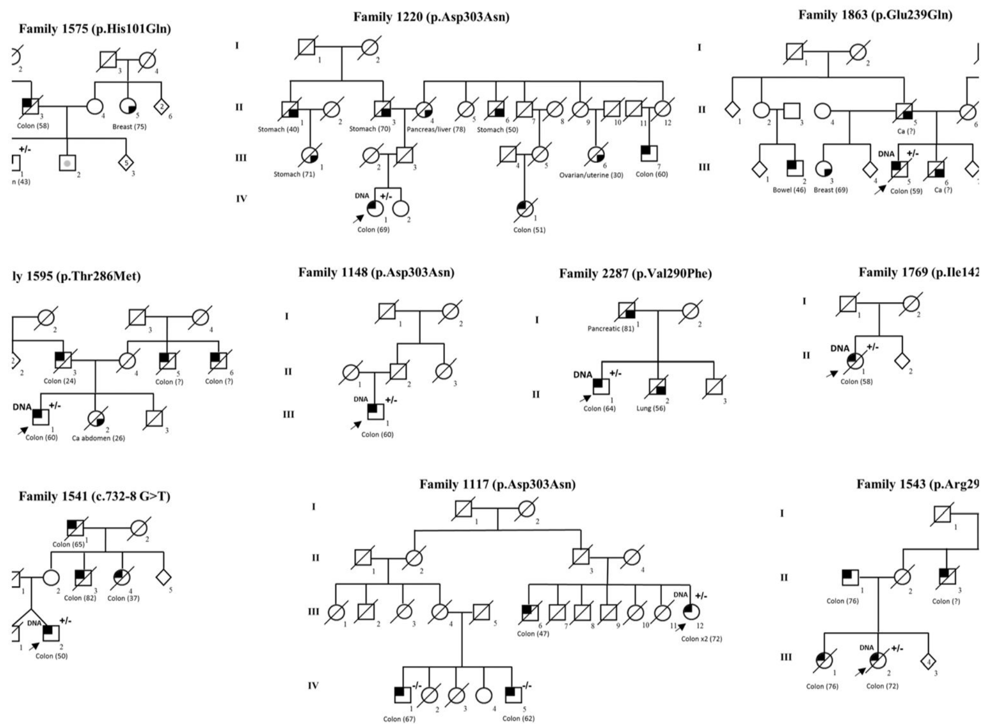


**Figure 1. Families from the Newfoundland Colorectal Cancer Registry with rare *GALNT12* variants**

Ten unrelated families harboring rare *GALNT12* alleles were identified in this study. Three families (1220, 1148, 1117) harbored the p.Asp303Asn allele, while the remaining variants were each present in probands of a single family. Eight families were MMR proficient and MSS, while two families (1575 and 1595) carried pathogenic MMR gene variants. Symbols: squares (males), circles (females), upper quadrant shaded (CRC), lower quadrant shaded (other cancer), grey shaded circle (polyps), diagonal line (deceased), arrow (proband), +/− (heterozygous variant). Age of cancer diagnoses are indicated in brackets. RefSeqs: NM_024642.4; NP_078918.3; NC_000009.12.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

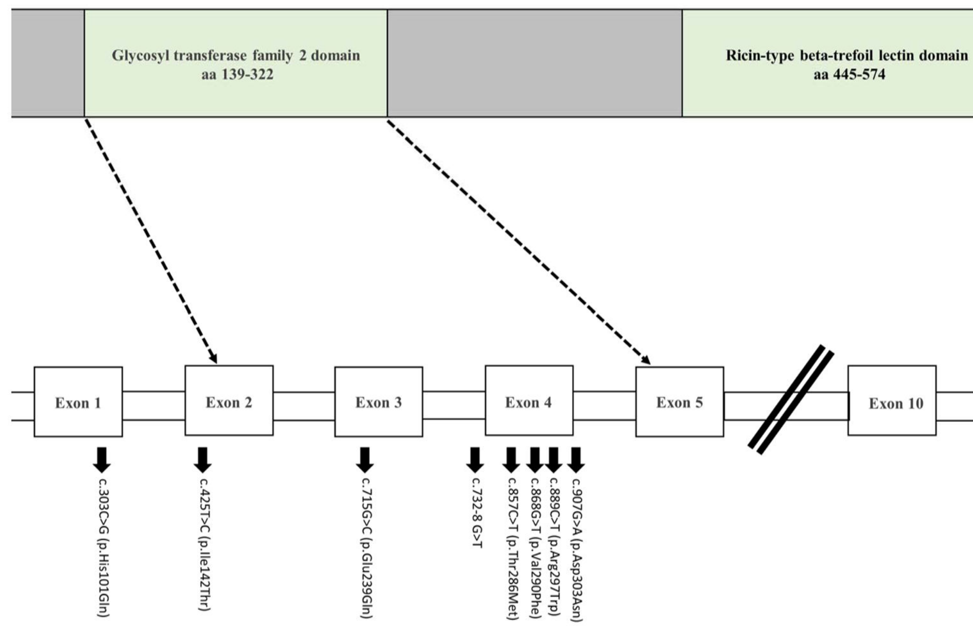


**Figure 2. Rare germline variants of *GALNT12* identified in CRC patients cluster within and around the glycosyltransferase domain of the protein**

Eight rare candidate *GALNT12* variants were identified in CRC cases in this study. Six out of the eight variants were situated within the glycosyltransferase domain, while two (p.His101Gln, p.Ile142Thr) were located upstream. RefSeqs: NM_024642.4; NP_078918.3; NC_000009.12.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

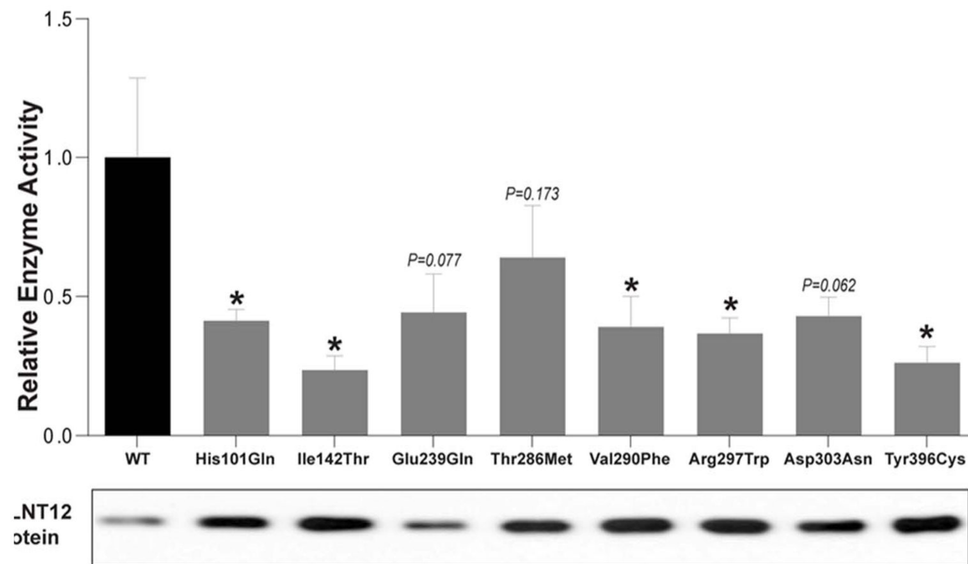


**Figure 3. Biochemical characterization of wild-type and mutant GALNT12 proteins**

Y-axis depicts the enzyme activities of mutant versions of GALNT12, relative to the wild-type (WT) protein (X-axis). Error bars represent standard error of the means derived from 3 independent replicate experiments. (*) indicates significant difference (P 0.05) in enzyme activity of respective mutant GALNT12 as compared to the WT protein, estimated by one-tailed Student's t-test. Shown below the graph is a representative Western blot image demonstrating protein expression of wild-type and mutant versions of GALNT12 in SW480 cells transfected with respective cDNA constructs. The p.Tyr396Cys GALNT12 variant was identified in our previous clinical cohort-based study (Clarke et al., 2012) and was included in this *in vitro* enzyme assay to test its enzymatic potential.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

**Table 1**

Rare *GALNT12* variants identified in a population-based cohort of incident CRC cases from Newfoundland.

| *GALNT12* Variant | Type | dbSNP rsID | NFCCR Cohort | GnomAD (MAF/AC/Total Alleles) | SIFT | Polyphen-2 HumVar | REVEL | CADD | NL Controls (CRC-free) |
|---|---|---|---|---|---|---|---|---|---|
| **c.303C>G (p.His101Gln)** | SNV | rs201926457 | 1/479 | 0.0001516/27/178098 | 0.01 (D) | 0.979 (PRD) | 0.3301 | 28.1 | 0/400 |
| **c.425T>C (p.Ile142Thr)** | SNV | rs757214097 | 1/479 | NA | 0 (D) | 0.998 (PRD) | 0.76 | 27.5 | 0/400 |
| **c.715G>C (p.Glu239Gln)** | SNV | rs777144221 | 1/479 | 0.00008728/17/194766 | 0.05 (D) | 0.99 (PRD) | 0.5701 | 29.6 | 0/400 |
| **g.98831764G>T (c.732-8 G>T)** | intronic/splice variant | rs763682300 | 1/479 | 0.000004061/1/246220 | NA | NA | NA | 12.18 | 0/400 |
| **c.857C>T (p.Thr286Met)** | SNV | rs548915885 | 1/479 | 0.00002844/7/246124 | 0.01 (D) | 0.999 (PRD) | 0.36 | 34 | 0/400 |
| **c.868G>T (p.Val290Phe)** | SNV | rs371949942 | 1/479 | 0.00009383/26/277090 | 0 (D) | 0.94 (PRD) | 0.36 | 25 | 0/400 |
| **c.889C>T (p.Arg297Trp)** | SNV | rs149726976 | 1/479 | 0.0002744/76/276924 | 0 (D) | 0.908 (POD) | 0.3000 | 27.3 | 1/400 |
| **c.907G>A (p.Asp303Asn)** | SNV | rs145236923 | 3/479 | 0.001244/344/276526 | 0.47 (T) | 0.999 (PRD) | 0.4000 | 24.8 | 1/400 |

Abbreviations: SNV (single nucleotide variant); nSNV (nonsynonymous single nucleotide variant); UTR (untranslated region); NFCCR (Newfoundland Colorectal Cancer Registry); NL (Newfoundland & Labrador); SIFT (sorting tolerant from intolerant); T (tolerated); D (damaging); B (benign); POD (possibly damaging); PRD (probably damaging); NA (not applicable); MAF (minor allele frequency); AC (allele count); dbSNP (the single nucleotide polymorphism database) (available at https://www.ncbi.nlm.nih.gov/SNP/) (Assembly GRCh38.p7); rsID (reference SNP cluster ID); GnomAD (Genome Aggregation Database) (available at http://gnomad.broadinstitute.org/) (version r2.0.2); CADD (Combined Annotation Dependent Depletion); REVEL (Rare Exome Variant Ensemble Learner). Note: RefSeqs used included NM_024642.4; NP_078918.3; NC_000009.12.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

## Table 2

Clinical and pathological characteristics of Newfoundland Colorectal Cancer Registry patients harboring rare candidate *GALNT12* variants.

| Patient | Family | Variant | Age of diagnosis | Tumor location | Microsatellite stability | MMR protein IHC | MMR variants | Risk criteria |
|---|---|---|---|---|---|---|---|---|
| 4III-1 | 1575 | c.303C>G; p.His101Gln | 43 | Rectum | MSI-H | MSH6 deficient | *MSH6*: c.3514dupA; Arg1172Lysfs*5 | Bethesda |
| II-1 | 1769 | c.425T>C; p.Ile142Thr | 58 | Transverse | MSS | Intact | Negative | NA |
| III-5 | 1863 | c.715G>C; p.Glu239Gln | 59 | Rectum | MSS | Intact | Negative | Bethesda |
| III-2 | 1541 | g.98831764G>T; c.732-8 G>T | 51 | Transverse | MSS | Intact | Negative | AC-1,FCCTX, ACMAC, Bethesda |
| III-1 | 1595 | c.857C>T; p.Thr286Met | 60 | Descending | MSI-H | MLH1 deficient | *MLH1*: c.793C>T; p.His264Leufs*2 | AC-1, ACMAC, Bethesda |
| II-1 | 2287 | c.868G>T; p.Val290Phe | 64 | Rectum | MSS | Intact | Negative | NA |
| III-2 | 1543 | c.889C>T; Arg297Trp | 72 | Ascending | MSS | Intact | Negative | NA |
| III-1 | 1148 | c.907G>A; p.Asp303Asn | 72 | Cecum | MSS | Intact | Negative | NA |
| IV-1 | 1220 | c.907G>A; p.Asp303Asn | 59 | Descending | MSS | Intact | Negative | NA |
| III-12 | 1148 | c.907G>A; p.Asp303Asn | 60 | Sigmoid | MSS | Intact | Negative | Bethesda |

High microsatellite instability (MSI-H), microsatellite stable (MSS), immunohistochemistry (IHC), mismatch repair (MMR). Risk Criteria: Familial colorectal cancer type X (FCCTX), Revised Bethesda Criteria (Bethesda), Age and Cancer Modified Amsterdam Criteria (ACMAC). RefSeqs used include NM_024642.4; NP_078918.3; NC_000009.12.