BRIEF REPORT

Human Mutation

OFFICIAL JOURNAL

HGVS
HUMAN GENOME
VARIATION SOCIETY
www.hgvs.org

GALNT12 is Not a Major Contributor of Familial Colorectal Cancer Type X

Nuria Seguí,[1] Marta Pineda,[1] Matilde Navarro,[1] Conxi Lázaro,[1] Joan Brunet,[2] Mar Infante,[3] Mercedes Durán,[3] José Luis Soto,[4] Ignacio Blanco,[1] Gabriel Capellá,[1] and Laura Valle[1]*

[1]Hereditary Cancer Program, Catalan Institute of Oncology, IDIBELL, Hospitalet de Llobregat, Spain; [2]Hereditary Cancer Program, Catalan Institute of Oncology, IDIBGi, Girona, Spain; [3]Instituto de Biología y Genética Molecular, IBGM-UVA-CSIC, Valladolid, Spain; [4]Elche University Hospital, Elche, Spain

ABSTRACT: Previous evidence indicates that mutations in the *GALNT12* gene might cause a fraction of the unexplained familial colorectal cancer (CRC) cases: *GALNT12* is located in 9q22-33, in close proximity to a CRC linkage peak; and germline missense variants that reduce the enzymatic activity of the protein have been identified in CRC patients, some of them with familial CRC history. We hypothesized that mutations in *GALNT12* might explain part of the high-risk families grouped as familial CRC type X (fCRC-X), that is, Amsterdam-positive families with mismatch repair proficient tumors. We sequenced the coding regions of the gene in 103 probands of fCRC-X families, finding no functionally relevant mutations. Our results rule out *GALNT12* as a major high CRC susceptibility gene. Additional studies are required to provide further evidence about its role as a moderate/low susceptibility gene in familial aggregation of cancer.

Hum Mutat 35:50–52, 2014. © 2013 Wiley Periodicals, Inc.

KEY WORDS: *GALNT12*; hereditary colorectal cancer; colorectal cancer susceptibility; penetrance

Family history is well established to be one of the strongest risk factors for the development of colorectal cancer (CRC) and it is thought to involve approximately 20% of all CRC cases. However, only a minority of all CRC cases (2%–6%) are explained by germline genetic mutations in well-known high penetrance genes [reviewed by Lynch et al., 2009]. The proportion of CRC families that fulfill the most stringent criteria for hereditary CRC, that is, the Amsterdam criteria, and do not show a mismatch repair (MMR) defect is high (~40%). These individuals are at increased risk of developing CRC and therefore require strict cancer surveillance strategy. The genetic cause of the CRC familial aggregation in these families is unknown and they have been grouped as familial CRC type X (fCRC-X) [reviewed by Ku et al., 2012]. Several dominantly act-

ing predisposition loci mapping to different chromosomal regions, such as 9q22.2-31.2, 5q14-q22, or 3q13.31-q27.1, have been identified through genome-wide linkage studies in CRC families, but so far no causal gene has been reported [Gray-McGuire et al., 2010; Picelli et al., 2008; Tomlinson et al., 1999].

Previous evidence indicates that mutations in the *GALNT12* gene (MIM #610290), which codes for the enzyme *N*-acetylgalactosaminyltransferase-type 12, might explain familial CRC cases of unknown etiology [Clarke et al., 2012; Guda et al., 2009]. This gene, whose protein product is involved in the O-glycosylation of mucin-type glycans, shows high expression levels in the normal colon and is downregulated in a significant proportion of colorectal tumors [Guo et al., 2002; 2004]. *GALNT12* is located in 9q21-33, in close proximity to the linkage peak in 9q22-31, recurrently found when studying familial CRC cases [Gray-McGuire et al., 2010; Kemp et al., 2006; Skoglund et al., 2006; Wiesner et al., 2003]. In 2009, Guda et al. identified functionally relevant germline *GALNT12* mutations in seven out of 272 (2.6%) colon cancer patients. Of note, no mutations were present in 192 cancer-free controls. Recently, Clarke et al. (2012) studied the occurrence of *GALNT12* mutations in a cohort of 118 CRC families of unknown genetic cause. They identified two missense mutations present in four different Bethesda-positive families. However, no mutations were detected in 26 probands that met the Amsterdam I criteria. Based on the previous evidence implicating *GALNT12* in CRC predisposition, we wanted to expand the study of the relevance of *GALNT12* mutations in the etiology of fCRC-X.

A total of 103 fCRC-X families (all Caucasian) were included in the study: forty-three had been referred to the Genetic Counseling Units of the Catalan Institute of Oncology in the Spanish region of Catalonia between 1998 and 2011; 27 families were recruited by the Hereditary Cancer Unit of the Instituto de Biología y Genética Molecular (IBGM) of the University of Valladolid and the Spanish Research Council (CSIC), between 2006 and 2011; and 33 came from the Hereditary Cancer Program of the Spanish region of Valencia, which belongs to the Valencian Biobank Network, and were collected between 2005 and 2012. All families fulfilled Amsterdam criteria and did not have MMR deficiency (either tumor microsatellite instability and/or lack of expression of the MMR proteins MLH1, MSH2, MSH6, or PMS2). Of all 103 families, 45.6% fulfilled Amsterdam I and 54.4% Amsterdam II. The mean age at cancer diagnosis was 48.4 for the sequenced probands. Informed consent was obtained from all subjects and the study received the approval of the Ethics Committee of IDIBELL (ref. PR073/12). Supp. Table S1 shows the characteristics of the families included in the study.

All exons and exon–intron boundaries were sequenced using a standard protocol for automated direct Sanger sequencing. Primer sequences and PCR conditions are available upon request.

Table 1. Germline *GALNT12* Variants Identified in 103 fCRC-X Index Patients

Nucleotide change[a]	Amino-acid change	refSNP	Location	n	Het/Hom	MAF	Population MAF[b] (dbSNP)	Population MAF[c] (ESP)	Reported in previous studies (MAF)
c.136G>A	p.G46R	rs10987768	Exon 1	1	1/0	0.005	0.09	–	Guda et al. (2009) (NA)
c.237C>T	p. =	–	Exon 1	1	1/0	0.005	–	–	
c.356A>T	p.E119V	rs1137654	Exon 1	18	18/0	0.087	–	0.058	Guda et al. (2009) (NA); Clarke et al. (2012) (0.08)
c.399T>C	p. =	–	Exon 2	1	1/0	0.011	–	–	
c.541+74G>T	–	rs73496150	Intron 2	2	2/0	0.010	0.022	–	
c.781G>A	p.D261N	rs41306504	Exon 4	2	2/0	0.010	0.008	0.008	Guda et al. (2009) (NA); Clarke et al. (2012) (0.015)
c.917+24C>T	–	rs41297187	Intron 4	4	3/1	0.024	0.043	0.017	
c.1036–42delT	–	rs3216734	Intron 5	31	25/6	0.180	0.151	0.163	
c.1036–4G>A	–	–	Intron 5	1	1/0	0.005	–	–	
c.1344+61G>T	–	rs3824516	Intron 7	12	11/1	0.063	0.058	–	
c.1392C>T	p. =	rs35616709	Exon 8	1	1/0	0.005	0.001	–	
c.1458+58G>T	–	rs1885608	Intron 8	81	44/37	0.573	0.230	–	
c.1497C>T	p. =	rs35632007	Exon 9	1	1/0	0.005	0.001	0.002	Clarke et al. (2012) (0.008)
c.1605+4G>A	–	rs79574929	Intron 9	1	1/0	0.005	0.006	0.012	
c.1707G>C	p. =	rs2273846	Exon 10	5	5/0	0.024	0.147	0.083	Clarke et al. (2012) (0.043)
c.*67A>T	–	rs78514784	3′-UTR	2	2/0	0.010	0.017	–	
c.*171A>G	–	rs2273847	3′-UTR	5	5/0	0.024	0.147	–	Clarke et al. (2012) (0.031)
c.*421G>A	–	rs2273848	3′-UTR	11	11/0	0.053	0.059	–	Clarke et al. (2012) (0.088)
c.*499T>A	–	–	3′-UTR	1	1/0	0.011	–	–	
c.*547dup	–	–	3′-UTR	1	1/0	0.011	–	–	

[a]RefSeq NM_024642.4.
[b]MAF reported at the dbSNP and 1000 Genomes databases (http://ncbi.nlm.nih.gov/projects/SNP/).
[c]MAF reported at the NHLBI Exome Sequencing Project (ESP) (http://evs.gs.washington.edu/EVS/). Most intronic changes are not covered.
Het, heterozygous; Hom, homozygous; MAF, minor allele frequency; NA, not available.

Table 2. In Silico Prediction of Functional Effects of the Identified Variants

Variant	refSNP	Location	Protein prediction (score)			Splice site (SS) prediction[a]			
			PolyPhen-2 (HumDiv / HumVar)	Condel	SIFT	Distance to the nearest SS (bp)	SS	WT score	Variant score
c.136G>A (p.G46R)	rs10987768	Exon 1	Benign (0.002) / benign (0.001)	Neutral (0.011)	Tolerated (0.18)	236	D	0.8	0.8
c.237C>T (p. =)	–	Exon 1	–	–	–	135	D	0.8	0.8
							D2	NR	0.52
c.356A>T (p.E119V)	rs1137654	Exon 1	Possibly damaging (0.791) / benign (0.198)	Neutral (0.011)	Tolerated (0.18)	16	D	0.8	0.8
c.399T>C (p. =)	–	Exon 2	–	–	–	28	A	0.89	0.89
						143	D	1	1
c.541+74G>T	rs73496150	Intron 2	–	–	–	74	D	1	1
c.781G>A (p.D261N)	rs41306504	Exon 4	Probably damaging (1)/probably damaging (0.968)	Neutral (0.449)	Tolerated (0.22)	50	A	0.56	0.56
						137	D	0.87	0.87
c.917+24C>T	rs41297187	Intron 4	–	–	–	24	D	0.87	0.87
c.1036–42delT	rs3216734	Intron 5	–	–	–	42	A	0.95	0.95
c.1036–4G>A	–	Intron 5	–	–	–	4	A	0.95	0.98
c.1344+61G>T	rs3824516	Intron 7	–	–	–	61	D	0.98	0.98
c.1392C>T (p. =)	rs35616709	Exon 8	–	–	–	48	A	0.88	0.88
						67	D	0.99	0.99
c.1458+58G>T	rs1885608	Intron 8	–	–	–	58	D	0.99	0.99
c.1497C>T (p. =)	rs35632007	Exon 9	–	–	–	42	A	0.8	0.8
						106	D	0.96	0.96
c.1605+4G>A	rs79574929	Intron 9	–	–	–	4	D	0.96	1
c.1707G>C (p. =)	rs2273846	Exon 10	–	–	–	102	A	0.98	0.98
c.*67A>T	rs78514784	3′-UTR	–	–	–	208	A	0.98	0.98
c.*171A>G	rs2273847	3′-UTR	–	–	–	312	A	0.98	0.98
c.*421G>A	rs2273848	3′-UTR	–	–	–	562	A	0.98	0.98
c.*499T>A	–	3′-UTR	–	–	–	640	A	0.98	0.98
c.*547dup	–	3′-UTR	–	–	–	684	A	0.98	0.98

[a]Prediction calculated by NNSplice 0.9.
For splice site prediction, major alterations represent creation or destruction of a splice site, or score modifications ≥45%.
SS, splice site; bp, base pairs; WT, wild-type; A, acceptor consensus splice site; D, donor consensus splice site; D2, new donor splice site; NR, splice site not recognized.

Sequencing was performed on an ABI Sequencer 3730 and data analyzed using Mutation Surveyor v. 3.10. Identified variants were submitted to the LOVD database (http://www.lovd.nl/GALNT12).

A total of 20 variants were identified in 90 patients. Of the 20 variants, eight were located in protein-coding regions, seven in introns, and five in the 3′-UTR. Three of the eight changes in protein-coding regions, c.136G>A (p.G46R), c.356A>T (p.E119V), and c.781G>A (p.D261N), were nonsynonymous. The change p.E119V was identified in 18 fCRC-X patients, p.D261N in 2, and p.G46R in 1 patient. The three variants had been previously described in public databases with population minor allele frequencies of 0.06, 0.008, and 0.09, respectively (Table 1). Also, previous studies identified these three variants both in cancer cases and controls [Clarke et al., 2012; Guda et al., 2009]. No relevant effects on the protein were predicted by the

in silico algorithms SIFT and Condel, while PolyPhen-2 predicted functional effects for p.D261N (score ~1) and p.E119V (score 0.8, when applying HumanDiv) (Table 2). Nevertheless, Guda et al. (2009) reported that none of these three variants altered the enzymatic activity of the protein.

The remaining five variants in protein-coding regions predicted to translate into synonymous amino-acid changes, affecting codons 79, 133, 464, 499, and 569. All but two, c.237C>T and c.399C>T, had been previously described in public databases (dbSNP or ESP). The population allelic frequencies reported for other two, c.1392C>T and c.1497C>T, were below 2% (Table 1). The in silico algorithm NNSplice [Reese et al., 1997] only predicted a possible effect on the splicing process for c.237C>T, where a new donor splice would be created (score 0.52) (Table 2). However, the generation of this new donor splice site was not predicted by other algorithms, such as NetGene2 [Hebsgaard et al., 1996] or SoftBerry [Burset et al., 2001]. Moreover, this variant did not segregate with the disease in the family: the father of the proband, diagnosed with rectal cancer at age 64, did not carry the variant, being therefore absent in the CRC-affected family branch.

Of the intronic variants identified, only one, c.1036–4G>A, had not been reported as a polymorphism. Likewise, the change c.1605+4G>A (rs79574929) had a population MAF of 0.006–0.012 (sources: dbSNP and ESP, respectively). Despite their proximity to the corresponding consensus splice sites, no relevant effect on the splicing process was predicted (Table 2). Regarding c.1036–4G>A, it was not present in the mother of the proband, who was diagnosed with two metachronous colon tumors at the age of 57 and 59 years, discarding it as the genetic cause of the CRC familial aggregation. Unfortunately, no cosegregation analysis could be performed for c.1605+4G>A. In summary, no functionally relevant variants were identified in any of the 103 fCRC-X families evaluated.

The results obtained in our series (*n* = 103 fCRC-X families), in line with those of Clarke et al., who found no mutations in 26 Amsterdam-positive families, support the notion that *GALNT12* is not a major high-penetrance gene for CRC predisposition. The lack of information on the familial cancer history of the carriers of *GALNT12* germline mutations in Guda et al. (2009) does not allow further investigation in this regard.

Considering the presence of *GALNT12* mutations in non-Amsterdam familial CRC cases, Clarke et al. identified two missense variants, c.907G>A (p.D303N) and c.1187A>G (p.Y396C) in 4 families that met the revised Bethesda criteria [Clarke et al., 2012]. The c.907G>A change, rs145236923, identified in three different families, partially inactivates the enzymatic activity of the protein, in contrast to the other mutations, germline and somatic, identified by Guda et al. (2009) which cause almost total inactivation of the enzyme. Moreover, the evidence of cosegregation of c.907G>A with CRC in one of the two families studied was weak. The second missense variant identified, c.1187A>G (p.Y396C), is not reported in public databases, is located in the catalytic domain of the protein, and in silico algorithms predict it to be functionally relevant [Clarke et al., 2012]. All reported patients carrying inactivating germline mutations developed CRC later in life (median age: 71), and half of them were diagnosed with multiple primary epithelial tumors, including breast and colon cancers [Guda et al., 2009]. Further studies are required to provide a definitive answer about the role of *GALNT12* mutations in this subset of familial cancer cases. Either germline mutations in *GALNT12*, as a moderate-risk gene, or in other genes located within the 9q22-31 linkage peak might still explain some familial CRC cases. Indeed, the strength of the studies that replicate the linkage signal in 9q22-31, which suggest that the disease

locus housed on 9q is specific to a familial syndrome with a phenotype of young age of onset and/or severity of colon neoplasia [Gray-McGuire et al., 2010], supports the idea that probably the other genes under the linkage peak, such as *ZNF367*, *HABP4*, and *GABBR2*, might be relevant in CRC susceptibility.

In conclusion, our findings indicate that *GALNT12* is not a major high penetrance gene for CRC. Further comprehensive studies in Amsterdam-negative CRC families are required to clarify the gene's role as moderate/low susceptibility gene in CRC familial aggregation.

References

Burset M, Seledtsov IA, Solovyev VV. 2001. SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res 29(1):255–259.

Clarke E, Green RC, Green JS, Mahoney K, Parfrey PS, Younghusband HB, Woods MO. 2012. Inherited deleterious variants in GALNT12 are associated with CRC susceptibility. Hum Mutat 33(7):1056–1058.

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, et al. 2010. Confirmation of linkage to and localization of familial colon cancer risk haplotype on chromosome 9q22. Cancer Res 70(13):5409–5418.

Guda K, Moinova H, He J, Jamison O, Ravi L, Natale L, Lutterbaugh J, Lawrence E, Lewis S, Willson JK, Lowe JB, Wiesner GL, et al. 2009. Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. Proc Natl Acad Sci USA 106(31):12921–12925.

Guo JM, Chen HL, Wang GM, Zhang YK, Narimatsu H. 2004. Expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase-12 in gastric and colonic cancer cell lines and in human colorectal cancer. Oncology 67(3–4): 271–276.

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(1–3):211–218.

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. 1996. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Res 24(17):3439–3452.

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, et al. 2006. Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. Cancer Res 66(10):5003–5006.

Ku CS, Cooper DN, Wu M, Roukos DH, Pawitan Y, Soong R, Iacopetta B. 2012. Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X. Mod Pathol 25(8):1055–1068.

Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. 2009. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet 76(1):1–18.

Picelli S, Vandrovcova J, Jones S, Djureinovic T, Skoglund J, Zhou XL, Velculescu VE, Vogelstein B, Lindblom A. 2008. Genome-wide linkage scan for colorectal cancer susceptibility genes supports linkage to chromosome 3q. BMC Cancer 8:87.

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol 4(3):311–323.

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(2):e7.

Tomlinson I, Rahman N, Frayling I, Mangion J, Barfoot R, Hamoudi R, Seal S, Northover J, Thomas HJ, Neale K, Hodgson S, Talbot I, et al. 1999. Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22. Gastroenterology 116(4):789–795.

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(22):12961–12965.