Gastroenterology 2017;152:75–77

Validation of Recently Proposed Colorectal Cancer Susceptibility Gene Variants in an Analysis of Families and Patients—a Systematic Review

Peter Broderick,[1,*] **Sara E. Dobbins**,[1,*] **Daniel Chubb**,[1,*] Ben Kinnersley,[1] Malcolm G. Dunlop,[2] Ian Tomlinson,[3] and Richard S. Houlston[1,4]

[1]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK; [2]Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh; [3]Molecular and Population Genetics Laboratory and NIHR Biomedical Research Centre, Oxford Centre for Cancer Gene Research, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; and [4]Division of Pathology, The Institute of Cancer Research, London, UK

See editorial on page 12.

High-throughput sequencing analysis has accelerated searches for genes associated with risk for colorectal cancer (CRC); germline mutations in *NTHL1, RPS20, FANCM, FAN1, TP53, BUB1, BUB3, LRP6*, and *PTPN12* have been recently proposed to increase CRC risk. We attempted to validate the association between variants in these genes and development of CRC in a systematic review of 11 publications, using sequence data from 863 familial CRC cases and 1604 individuals without CRC (controls). All cases were diagnosed at an age of 55 years or younger and did not carry mutations in an established CRC predisposition gene. We found sufficient evidence for *NTHL1* to be considered a CRC predisposition gene—members of 3 unrelated Dutch families were homozygous for inactivating p.Gln90Ter mutations; a Canadian woman with polyposis, CRC, and multiple tumors was reported to be heterozygous for the inactivating NTHL1 p.Gln90Ter/c.709+1G>A mutations; and a man with polyposis was reported to carry p.Gln90Ter/p.Gln287Ter; whereas no inactivating homozygous or compound heterozygous mutations were detected in controls. Variants that disrupted *RPS20* were detected in a Finnish family with early-onset CRC (p.Val50SerfsTer23), a 39-year old individual with metachronous CRC (p.Leu61GlufsTer11 mutation), and a 41-year-old individual with CRC (missense p.Val54Leu), but not in controls. We therefore found published evidence to support the association between variants in *NTHL1* and *RPS20* with CRC, but not of other recently reported CRC susceptibility variants. We urge the research community to adopt rigorous statistical and biological approaches coupled with independent replication before making claims of pathogenicity.

Keywords: Colon Cancer; Inherited; Germline; Exome Sequencing.

Understanding the genetics of familial colorectal cancer (CRC) is clinically important to discriminate between high- and low-risk groups. Mutations in 11 genes are well-established to confer significant increases in CRC risk and testing for these is common in clinical practice. Despite this in many CRC families no genetic diagnosis can be made. While the availability of high-throughput sequencing has accelerated searches for new CRC genes, there are challenges in assigning pathogenicity to identified variants.

Here we reviewed the data supporting recent assertions that *NTHL1, RPS20, FANCM, FAN1, TP53, BUB1, BUB3, LRP6,* and *PTPN12* are CRC susceptibility genes using an evidence-based framework (Supplementary Material).[1–7] To search for independent evidence of a role in CRC risk we analyzed sequencing data on 863 familial CRC cases and 1604 controls.[8] All cases were diagnosed aged ≤55 years and were mutation-negative for known CRC genes.

Evidence for variation in *NTHL1,* which like *MUTYH* performs base-excision repair, as a cause of recessive CRC has been provided by 3 unrelated Dutch families homozygous for the rare inactivating p.Gln90Ter mutation (Supplementary Material, Supplementary Table 1).[6] The tumor mutation spectrum was enriched for C>T transitions, consistent with defective base-excision repair. Subsequently compound heterozygosity for inactivating *NTHL1* p.Gln90Ter/c.709+1G>A mutations was identified in a Canadian woman diagnosed with polyposis, CRC, and multiple tumors.[9] Tumors were again enriched for somatic C>T transitions. While we found no p.Gln90Ter homozygotes among our whole-exome sequencing (WES) cases, a 41-year old male case with coincident polyposis harbored p.Gln90Ter/p.Gln287Ter. No inactivating homozygotes or compound heterozygotes were seen among our 1604 controls.

WES of a Finnish Amsterdam-positive family demonstrated significant segregation of *RPS20* p.Val50SerfsTer23 with early-onset CRC (logarithm of odds score=3.0;

*Authors share co-first authorship.

Abbreviations used in this paper: CRC, colorectal cancer; MAF, minor alelle frequency; WES, whole-exome sequencing.

Most current article

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2016.09.041

Supplementary Material, Supplementary Table 1).[3] No disruptive *RPS20* variants have been catalogued by the Exome-Aggregation Consortium, which contains WES data for 60,706 individuals of diverse ancestries,[10] suggesting the gene is intolerant to mutation. Therefore, it is notable that in our WES series, we identified the disruptive p.Leu61GlufsTer11 mutation in a 39-year-old with metachronous CRC. Furthermore, we identified the deleterious missense p.Val54Leu in an Amsterdam-positive 41-year-old case. No rare missense/disruptive mutations identified in the 1604 controls.

Smith et al[5] identified *FANCM* p.Arg1931Ter in 2 sporadic CRC cases with cancers showing loss of the wild-type allele (loss of heterozygosity)[5]. p.Arg1931Ter has been shown to induce exon skipping resulting in decreased DNA repair (Supplementary Material, Supplementary Table 1). In our WES series, we detected p.Arg1931Ter in 4 cases and 1 control ($P = .02$; Supplementary Table 3). To seek further evidence for an association between p.Arg1931Ter and CRC, we investigated the frequency of this specific variant in 2 additional UK series totaling 5552 cases and 6792 population controls (published Illumina-Exome-BeadChip data[11]; Supplementary Material). Combining these data provided no evidence for an association (meta-analysis $P = .22$; Supplementary Figure 1).

FAN1 mutations have been reported as a cause of CRC in Amsterdam-positive families,[4] but evidence for segregation was weak ($P = .125$) and the evidence for any functional effect of mutation was only shown in noncolonic tissue (Supplementary Material, Supplementary Table 1). In our WES series, we found no significant increase in the burden of *FAN1* mutations in cases (Table 1; Supplementary Tables 2 and 3).

Germline mutation of *TP53,* archetypically associated with Li-Fraumeni syndrome, has recently been suggested to cause familial CRC at a frequency comparable to *APC.*[7] The assertion was, however, based on the flawed assumption that all rare missense changes seen were disease-causing with no consideration of mutation burden in controls (Supplementary Material, Supplementary Table 1). In our data no over-representation of *TP53* mutation was seen in cases (Table 1, Supplementary Tables 2 and 3).

By WES small numbers of early-onset CRC, *BUB1, BUB3, LRP6*, and *PTPN12* have been proposed as CRC predisposition genes.[1,2] The published evidence to support assertions is minimal (Supplementary Material, Supplementary Table 1) with no evidence of segregation or loss of heterozygosity. In addition, of the 2 *BUB1* mutation carriers, 1 also carried an *MLH1* mutation which, unlike *BUB1,* segregated with colorectal tumors. Only for *PTPN12* did the authors demonstrate an increase in the burden of mutation in cases vs controls ($P = .039$; Supplementary Material). Although we also observed an enrichment of missense *PTPN12* mutation in our WES cases ($P = .039$; Table 1, Supplementary Table 3), in light of the number of genes investigated, the evidence for a role in CRC predisposition remains weak.

In conclusion, a role for *NTHL1* as a bona fide CRC gene is supported by multiple lines of evidence. While compelling, the assertion that mutation of *RPS20* causes CRC remains to be established, as this observation is based on a single family and the mechanism by which ribosomal proteins might predispose to CRC is unclear. In contrast, evidence to support other genes as risk factors is currently lacking.

Investigators must remember that private variants are common; of the 7,404,909 variants listed in Exome-Aggregation Consortium, 54% are observed only once,[10] therefore, novel variants should be considered benign until proved otherwise. A studies power to detect a statistically significant association with any rare variant is typically weak, therefore, additional evidence must be considered including segregation of the genotype with disease in

Table 1. Gene Burden Analysis

		Disruptive mutations (stop-gain, frameshift)			Damaging mutations (disruptive, predicted-damaging, splice acceptor/donors)			All coding nonsynonymous variants		
Gene	Previously reported	Cases	Control	P_{Fisher}	Cases	Control	P_{Fisher}	Cases	Control	P_{Fisher}
BUB1	Disruptive	0	4	.31	1	8	.17	18	30	.76
BUB3	Missense	0	2	.55	0	4	.31	1	5	.67
FAN1	Disruptive/missense	0	2	.55	15	17	.19	32	45[a]	.23
FANCM	Disruptive/missense	5	1	**.02**	23	33	.33	51[b]	67[b]	.06
LRP6 (BPD[c])	Missense	0	0	—	6 (4)	17 (13)	.51 (.45)	17 (8)	37 (21)	.67
PTPN12	Missense	0	1	1.00	6	5	.21	12	9	**.04**
RPS20	Disruptive	1	0	.35	2	0	.12	2	0	.12
TP53	Missense	1	0	.35	1	1	1.00	1	4	.66

NOTE. Number of cases (n = 863) and controls (n = 1604) with rare (MAF <1%) mutations in postulated CRC genes. *P* values calculated using Fisher's exact test, *P* values <.05 are in bold.
BPD, β-propellor domain.
[a]Total number of variants in controls = 46; 1 sample has 2 *FAN1* missense.
[b]Totals number of variants in cases = 52, in controls = 69; 3 samples have 2 *FANCM* missense.
[c]Number of variants within BPD. All 3 variants identified by de Voer et al[2] were within BPD.