

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

Peter Broderick<sup>1\*</sup>, Sara E Dobbins<sup>1\*</sup>, Daniel Chubb<sup>1\*</sup>, Ben Kinnersley<sup>1</sup>, Malcolm G Dunlop<sup>2</sup>, Ian Tomlinson<sup>3</sup>, Richard S Houlston<sup>1,4</sup>

1. Division of Genetics and Epidemiology, The Institute of Cancer Research, London SM2 5NG, UK.
2. Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh EH4 2XU
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 OX3 7BN, UK.
4. Division of Pathology, The Institute of Cancer Research, London SM2 5NG, UK

\* These authors contributed equally to this work.

Correspondence to: Richard S Houlston, E-mail: richard.houlston@icr.ac.uk; Tel: +44 (0) 208 722 4175; Fax: +44 (0) 722 4365.

### **CONFLICT OF INTEREST**

Peter Broderick, Sara E Dobbins, Daniel Chubb, Ben Kinnersley, Malcolm G Dunlop, Ian Tomlinson, Richard S Houlston: None to declare

### **ACKNOWLEDGEMENTS**

This work was supported by Cancer Research UK Research (C1298/A8362, Bobby Moore Fund for Cancer Research UK) and the European Union (FP7/2007-2013) under Grant No. 258236, FP7 collaborative project SYSCOL. D.C. was funded by a grant from Bloodwise. Additional support was

provided by the National Cancer Research Network and the National Health Service (NHS). In Oxford, the work was funded by the Oxford Comprehensive Biomedical Research Centre core infrastructure support to the Wellcome Trust Centre for Human Genetics, Oxford (Wellcome Trust 090532/Z/09/Z). In Scotland, the work was funded by a Cancer Research UK (C348/A12076) and Medical Research Council Grant (MR/KO18647/1). This study makes use of the ICR1000 UK exome series data generated by Professor Nazneen Rahman's team at The Institute of Cancer Research, London. This work made use of samples generated by the 1958 Birth Cohort. Access to these resources was enabled via the 58READIE Project funded by Wellcome Trust and Medical Research Council (grant numbers WT095219MA and G1001799). This publication is supported by COST Action BM1206.

## **AUTHOR CONTRIBUTIONS**

Conception and design: Peter Broderick, Daniel Chubb, Sara E Dobbins, Richard S. Houlston

Collection and assembly of data: Peter Broderick, Daniel Chubb, Sara E. Dobbins, Malcolm G Dunlop, Ian Tomlinson, Richard S. Houlston

Data analysis and interpretation: Peter Broderick, Daniel Chubb, Ben Kinnersley, Sara E. Dobbins, Richard S. Houlston

Manuscript writing: All authors

Final approval of manuscript: All authors

## ABSTRACT

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.Leu61Glu fsTer11 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**

## ARTICLE

Understanding the genetics of familial CRC is clinically important to discriminate between high- and low-risk groups. Mutations in eleven 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)<sup>1-7</sup>. To search for independent evidence of a role in CRC risk we analyzed sequencing data on 863 familial CRC cases and 1,604 controls<sup>8</sup>. All cases were diagnosed aged  $\leq 55$  and were mutation-negative for known CRC genes.

Evidence for variation in *NTHL1*, which like *MUTYH* performs base-excision-repair (BER), as a cause of recessive-CRC has been provided by three unrelated Dutch families homozygous for the rare inactivating p.Gln90Ter mutation (Supplementary-Material, Supplementary-Table 1)<sup>6</sup>. The tumor mutation spectrum was enriched for C>T transitions, consistent with defective BER. 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<sup>9</sup>. Tumors were again enriched for somatic C>T transitions. While we found no p.Gln90Ter homozygotes amongst our WES cases, a 41-year old male case with co-incident polyposis harbored p.Gln90Ter/p.Gln287Ter. No inactivating homozygotes or compound heterozygotes were seen among our 1,604 controls.

Whole-exome sequencing (WES) of a Finnish Amsterdam-positive family demonstrated significant segregation of *RPS20* p.Val50SerfsTer23 with early-onset CRC (LOD score=3.0; Supplementary-Material, Supplementary-Table 1)<sup>3</sup>. No disruptive *RPS20* variants have been catalogued by the Exome-Aggregation-Consortium (ExAC), which contains WES data for 60,706 individuals of diverse ancestries<sup>10</sup> suggesting the gene is intolerant to mutation. Hence, 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 1,604 controls.

Smith *et al.* identified *FANCM* p.Arg1931Ter in two sporadic CRC cases with cancers showing loss of the wild-type allele (LOH)<sup>5</sup>. 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 four cases and one control ( $P=0.02$ ; Supplementary-Table 3). To seek further evidence for an association between p.Arg1931Ter and CRC, we investigated the frequency of this specific variant in two additional UK series totaling 5,552 cases and 6,792 population controls (published Illumina-Exome-BeadChip data<sup>11</sup>; Supplementary-Material). Combining these data provided no evidence for an association (Meta-analysis  $P=0.22$ ; Supplementary Figure 1).

*FAN1* mutations have been reported as a cause of CRC in Amsterdam-positive families<sup>4</sup>, but evidence for segregation was weak ( $P=0.125$ ) and the evidence for any functional effect of mutation was only shown in non-colonic 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&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*<sup>7</sup>. 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&3).

By WES small numbers of early-onset CRC, *BUB1*, *BUB3*, *LRP6* and *PTPN12* have been proposed as CRC predisposition genes<sup>1,2</sup>. The published evidence to support assertions is minimal (Supplementary-Material, Supplementary-Table 1) with no evidence of segregation or LOH. Moreover, of the two *BUB1* mutation carriers, one also carried a *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 versus controls ( $P=0.039$ ; Supplementary-Material). While we also observed an enrichment of missense *PTPN12* mutation in our WES cases ( $P=0.039$ ; 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 ExAC, 54% are observed only once<sup>10</sup>, 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 families, somatic mutation and functional studies with relevance to CRC biology. Critically, where multiple variants are considered within a gene, the burden of variation within controls must also be considered. Since the frequency of variants can be highly population-specific it is essential that controls used for comparison are well matched.

While there is a strong rationale for seeking to identify new CRC genes, well powered studies are required to mitigate against erroneous findings being asserted as causative and subsequently included in databases from which they are seldom deleted. The WES data we have generated represents the largest cohort of CRC exomes sequenced to date. The use of this dataset, which is publically available, to validate observations from small sequencing studies should act to limit the reporting of false positive results. Finally, the evidence framework we have implemented to assess the validity of proposed CRC genes, provides a robust strategy for establishing clinically actionable genes.

## TABLES AND FIGURES

**Table 1: Gene Burden analysis.** Number of cases (n=863) and controls (n=1,604) with rare (MAF<1%) mutations in postulated CRC genes. *P*-values calculated using Fishers exact test, *P*-values <0.05 are emboldened.

## REFERENCES

1. **de Voer RM, Geurts van Kessel A, Weren RD, Ligtenberg MJ**, et al. *Gastroenterology* 2013;145:544-7.
2. **de Voer RM, Hahn MM**, et al. *PLoS Genet* 2016;12:e1005880.
3. Nieminen TT, et al. *Gastroenterology* 2014;147:595-598 e5.
4. **Segui N, Mina LB**, et al. *Gastroenterology* 2015;149:563-6.
5. Smith CG, et al. *Hum Mutat* 2013;34:1026-34.
6. Weren RD, et al. *Nat Genet* 2015;47:668-71.
7. Yurgelun MB, et al. *JAMA Oncol* 2015;1:214-21.
8. **Chubb D, Broderick P, Dobbins SE**, et al. *Nat Commun* 2016;7:11883.
9. Rivera B, et al. *N Engl J Med* 2015;373:1985-6.
10. Lek M, et al. *Nature* 2016;536:285-91.
11. **Timofeeva MN, Kinnersley B**, et al. *Sci Rep* 2015;5:16286.

---

Author names in bold designate shared co-first authorship