

Data Supplementary (S)

Methods

Family Data

The proband (III:1) was recruited to the SAYO CRC Registry, as previously reported¹ and gave informed consent in writing. Further investigations were conducted clinically on the proband's mother (II:2). The study was conducted under ethics approval HREC/14/TQEHLMH/194 (The Queen Elizabeth Hospital, CALHN Office for Research, Adelaide, South Australia). Blood was sampled from both patients and whole exome sequencing (WES) of leucocyte DNA was performed for the proband as previously reported.¹ Subsequent sanger sequencing was performed on II:2. Medical records were audited for tumour-based testing results (MMR deficiency (dMMR) and somatic mutation testing). Next generation sequencing of *RNF43* and allelic loss studies in tumour DNA from both patients were performed to identify second hit mutations.

Whole Exome Sequencing Data Filtration and Interpretation

The WES data of III:1 underwent a comprehensive analysis of 133 genes that have been associated with cancer/CRC predisposition.²⁻⁴ Data filtration and interpretation were carried out using the “Varvis” exome analysis pipeline. The data were filtered for high-quality (coverage of >6 reads, a minimum quality score of 10, the alternative allele frequency of $\geq 25\%$), and rare variants (minor allele frequency (MAF) <1%) variants. The MAFs were taken from public databases (GnomAD, dbSNP) and an in-house database⁵. The American College of Medical Genetics and Genomics guidelines (ACMG)⁶ were used to evaluate sequence variants. Pathogenicity of putative germline variants were evaluated using classifications from public databases (ClinGen, ClinVar, InSiGHT, PubMed, OMIM, and HGMD). In silico prediction tools [SIFT and Provean (http://provean.jcvi.org/genome_submit_2.php?species=human), CADD (<https://cadd.gs.wash-ington.edu/>), PolyPhene-2 (<http://genetics.bwh.harvard.edu/pph2/>), and Mutation Taster (<http://mutationtaster.org/>)] were used for predicting the sequence variants for their protein deleteriousness potential. Pfam server (<http://pfam.xfam.org/>) was used to analyse protein sequences and identify protein domains. NetGene 2 (<http://www.cbs.dtu.dk/services/NetGene2/>), Alternative Splice Site Predictor (<http://wangcomputing.com/assp/index.html>), and Fruit Fly Splice Predictor (https://www.fruitfly.org/seq_tools/splice.html) servers were used for predicting the potential

splice site alterations. The Integrative Genomic Viewer (IGV) tool (IGV2.8.10) was utilized for the visualisation of sequence data and variants calls and Sanger sequencing was performed to confirm potentially damaging variants. Mutalyzer software (<https://mutalyzer.nl/>) was used to check the variant nomenclature.

Splicing Study of *RNF43*:c.375+1G>A

For cDNA analysis of the *RNF43* transcript, RNA was extracted from blood and mRNA was transcribed into cDNA by performing reverse transcriptase PCR and using RevertAid First Strand cDNA synthesis kit (Thermo Fisher Scientific Inc., Waltham, MA, USA). Exon and intron specific PCR primers were employed to investigate the *RNF43*:c.375+1G>A variant by PCR amplification of the relevant *RNF43* transcript region with the purpose to generate a product which a) span the wild-type normal splice boundaries of the *RNF43* exons which include exon 2 ending at coding nucleotide c.252, exon 3 spanning coding nucleotides c.253 to c.375, exon 4 spanning coding nucleotides c.376 to c.450, and exon 5 beginning at coding nucleotide c.451 and b) identify abnormal splicing which may occur within the amplified region including the use of allele-specific PCR primers targeted to *RNF43*:c.375+1G>A, and intronic primers downstream of exon 3. Primer sequences are available on request. PCR products were analyzed using bi-directional Sanger sequencing.

Mismatch Repair Deficiency Testing and Somatic Variants

Medical records were audited for tumour-based testing results. Mismatch repair deficiency (dMMR) was determined by immunohistochemical analysis⁷. MLH1 methylation analysis was performed for tumours exhibiting a loss of MLH1 or PMS2 staining. Somatic mutation testing was performed on DNA isolated from paraffin-embedded tumour tissue. Previously reported clinically significant variants in the *BRAF*, *EGFR*, *KIT*, *KRAS*, and *NRAS* genes were screened using a commercial panel (OncoFOCUS™)⁸. Next-generation sequencing of *RNF43* was performed using a custom IDT design (GMPfocus v2) sequenced on the Illumina NextSeq Sequencing System. Sequences were aligned to the human reference genome (hg19) using Burrows-Wheeler Aligner (BWA-mem). Variant calling was performed using the Genome Analysis Toolkit (GATK). GATK and Freebayes were used to analyse tumour samples. Variant annotation was performed using VariantGrid v2 and Alamut v2.11.

Extended results

Clinical Data

The proband (III:1) was diagnosed with a splenic flexure CRC, tubular adenoma (7mm; low-grade dysplasia), and ovarian mass at the age of 50 years. Histology confirmed a mucinous adenocarcinoma of the colon (120 x 85 x 30 mm) and a benign ovarian serous cystadenofibroma. II:2 was subsequently diagnosed with high grade, mucinous adenocarcinoma of the colon at the age of 65 years; and 7 polyps [2 hyperplastic (7 mm in sigmoid colon; 11 mm in ascending colon), 3 sessile serrated polyps (10 mm in sigmoid colon; 13 mm in ascending colon; unknown size in sigmoid colon); 2 tubular adenomas (30 mm with high-grade dysplasia in ascending colon; 15 mm with low-grade dysplasia in descending colon)]. The proband's maternal grandmother (I:3) was diagnosed with papillary invasive rectal adenocarcinoma of the colon at 72 years of age, but the polyp status was unknown. The proband's maternal half-uncle (II:1) was diagnosed with prostate cancer at the age of 62 years.

RNF43:c.375+1G>A Germline Variant

The WES data of III:1 was analyzed for 26 genes⁹ which have recently been recommended by the Collaborative Group of the Americas for Inherited Gastrointestinal Cancer to be used for evaluation of hereditary CRC². After applying a variant prioritization strategy to the WES data, no pathogenic/likely-pathogenic variant was identified. The data were screened using a further 107 genes present in commercial multigene panels^{10,11} and/or have been reported in the literature to be associated with hereditary CRC or cancer^{3,4,9,12}. Variant *RNF43*:c.375+1G>A was identified in the proband. This variant consists of a G>A nucleotide substitution at the +1 position of intron 3 of *RNF43*. In silico splice analysis tools predicted this variant to disrupt mRNA splicing and expected to result in an absent or disrupted protein product. The CADD score for this variant is 34 predicting its high potential for deleterious effects on the resulting protein. Exon 3 was present in the biologically relevant transcripts. ClinVar has no entries for this variant. This variant is neither found in population databases (GnomAD) nor in the Leiden Open Variation Database. This variant was confirmed in III:1 and II:2 by Sanger Sequencing (**Figure 2**).

Agarose gel electrophoresis demonstrated a PCR product from both wild-type control, III:1 and II:2 patient samples consistent in size with a transcript resulting from normal splicing. Additionally, in the III:1 and II:2 patient samples, a weakly amplified larger product was also observed.

DNA sequencing of the amplified products from the patient sample confirmed the presence of wild-type transcript (*RNF43*:r=) produced by normal splicing from the wild-type allele (*RNF43*:c.=). Notable, however, a weakly amplified product of the same size, but just above the limit of detection (<10% total mRNA), was also observed corresponding to the allele carrying *RNF43*:c.375+1G>A (**Figure S1**).

Next, we investigated if the weakly amplified *RNF43*:c.375+1G>A product was the consequence of allele-specific nonsense-mediated decay, and/or due to limitations of the assay to detect longer transcripts resulting from PCR bias in favour of shorter products. Allele-specific-PCR identified a single transcript, not observed in control samples, containing an insertion of 53 nucleotides from intron 3, spliced to the first nucleotide of exon 4. This product almost certainly results from preferential activation of a cryptic splice donor site beginning at position *RNF43*:c.375+54 effected by the *RNF43*:c.375+1G>A sequence change; the transcript is therefore described as NC_000017.10(NM_017763.5):r.375_376ins375+1_375+53, and predicted to encode *RNF43*:p.Ala126Ilefs*50 (**Figure 3, Figure S2**).

Bias amplification by allele-specific PCR of this shorter product precluded the detection of longer transcripts. However, downstream stepwise positioning of PCR primers within intron 3 enabled detection of an additional separate transcript containing an intronic insertion of greater than 226 nucleotides. While limitations of our assay prevent the precise characterisation of the cryptic splice donor site expected for this longer transcript, the resulting transcript may be described using position interval uncertainty, as NC_000017.10(NM_017763.5):r.375_376ins375+1_(375+226_376-1). This transcript is predicted to encode an aberrant protein (*RNF43*:p.Ala126Ilefs*50). Thus, independent of the exact position of the cryptic donor site activated downstream of position 375+226, a premature termination codon is encoded within the characterised region of the retained intron 3 sequence (226 nucleotides) (**Figure 3, Figure S3**).

These results demonstrate *RNF43*:c.375+1G>A effects almost complete loss of the normal wild type transcript from this allele in this family. It must be noted that the longer aberrant transcript (NC_000017.10(NM_017763.5):r.375_376ins375+1_(375+226_376-1) is also observed in the wild type allele (and normal control) (**Figure 3, Figure S3**). However, our findings indicate that this product is weakly expressed, comprising less than 5% of the total *RNF43* transcript. Therefore, *RNF43*:c.375+1G>A is responsible for abrogating normal splicing while concomitantly favouring at least one downstream cryptic splice donor site effecting abrogation

of normal splicing on the allele in which it resides. That normal splicing was absent and the resulting transcripts effect a frameshift insertion, identified *RNF43*:c.375+1G>A as equivalent to a constitutional germline variant in which a premature termination codon is encoded (**Figure 3**). According to the ACMG guidelines⁶, *RNF43*:c.375+1G>A was therefore classified as a likely-pathogenic variant (PVS1_Moderate, PM2, PS3, PP4).

Figure S

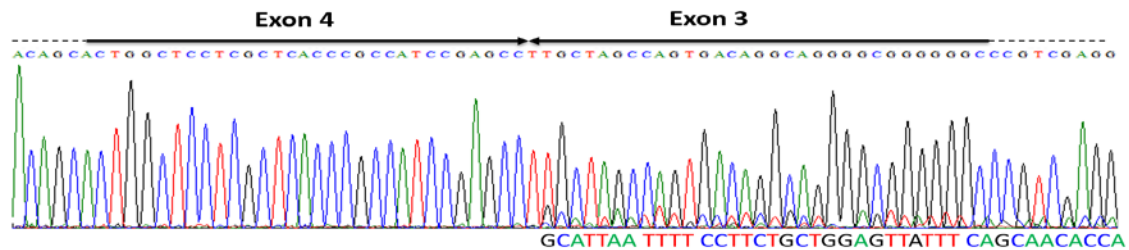


Figure S1. DNA sequence analysis of the amplified products from the patient sample confirmed the presence of wild-type transcript from the wild-type allele (*RNF43*:c.=) and a weakly amplified product from allele carrying *RNF43*:c.375+1G>A.

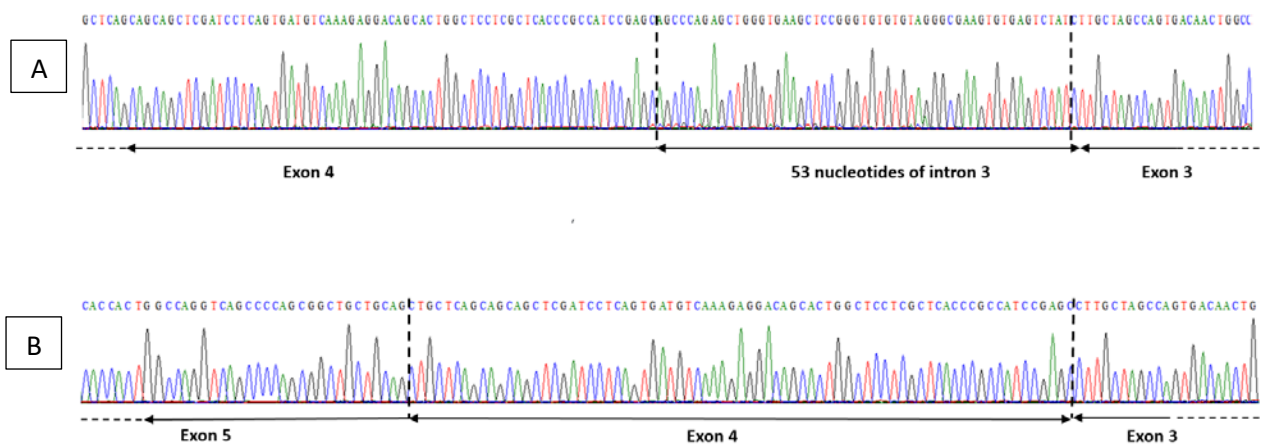


Figure S2. *RNF43* variant (A) and wild-type (B) allele-specific RT-PCR amplification and sequencing.

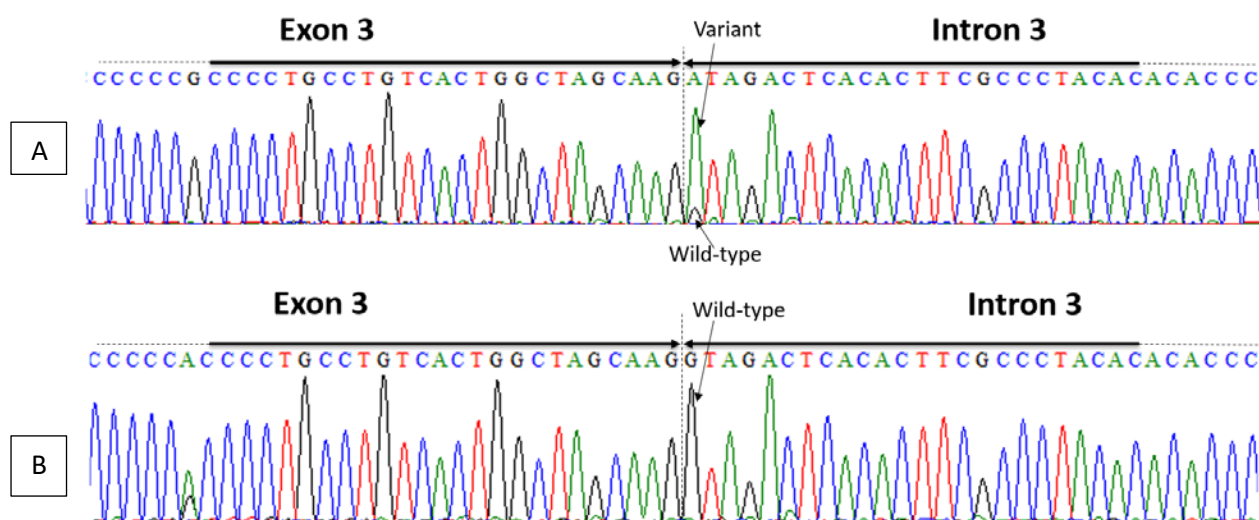


Figure S3. Downstream stepwise positioning of PCR primers within intron 3 in the patient (A) and the control (B) showing detection of an additional separate transcript containing an intronic insertion of greater than 226 nucleotides.

Table S1. Mismatch Repair Deficiency Testing and Identified Variants in Tumour.

Patient	Tissue	<i>BRAF</i> ^{V600E}	IHC	<i>RNF43</i> *	Allele fraction	Tumour Purity
III:1	CRC	mutated	MMR proficient	c.375+1G>A	61%	50%
II:2	CRC	mutated	MMR proficient	c.375+1G>A c.433C>T; p.(Arg145*)	50% 25%	60%

References

1. Mikaeel RR, Symonds EL, Kimber J, et al. Young-onset colorectal cancer is associated with a personal history of type 2 diabetes. *Asia-Pacific Journal of Clinical Oncology*. 2020.
2. Heald B, Hampel H, Church J, et al. Collaborative Group of the Americas on Inherited Gastrointestinal Cancer Position statement on multigene panel testing for patients with colorectal cancer and/or polyposis. *Familial cancer*. 2020;19(3):223-239.
3. Valle L, de Voer RM, Goldberg Y, et al. Update on genetic predisposition to colorectal cancer and polyposis. *Molecular aspects of medicine*. 2019.
4. Stoffel EM, Chittenden A. Genetic testing for hereditary colorectal cancer: challenges in identifying, counseling, and managing high-risk patients. *Gastroenterology*. 2010;139(5):1436-1441, 1441.e1431.
5. Hanses U, Kleinsorge M, Roos L, et al. Intronic CRISPR Repair in a Preclinical Model of Noonan Syndrome—Associated Cardiomyopathy. *Circulation*. 2020;142(11):1059-1076.
6. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics*. 2015;17(5):405-424.

7. Lindor NM, Burgart LJ, Leontovich O, et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2002;20(4):1043-1048.
8. OncoFOCUS™ Panel. Identify Actionable Somatic Mutations in Solid Tumors
2016; https://agenabio.com/wp-content/uploads/2016/11/51-20098R1.0_OncoFOCUS_Panel3_Flyer_1116_WEB.pdf. Accessed 25/02/2021.
9. Stoffel EM, Koeppe E, Everett J, et al. Germline Genetic Features of Young Individuals With Colorectal Cancer. *Gastroenterology*. 2018;154(4):897-905.e891.
10. Gene Dx. HEREDITARY CANCER REFERENCE LIST. 2018; https://www.genedx.com/wp-content/uploads/2018/01/40372_Hereditary-Cancer-Gene-List_Jan2018.pdf. Accessed 29/01/2021.
11. Sema 4. Sema4 Signal™ Hereditary Cancer Panel Guide. 2020; <https://cdn1.sema4.com/wp-content/uploads/Sema4-Signal-Hereditary-Cancer-Panel-Guide-1.pdf>. Accessed 29/01/2021.
12. Pearlman R, Frankel WL, Swanson B, et al. Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. *JAMA oncology*. 2017;3(4):464-471.