Accepted Manuscript

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

Peter Broderick, Sara E. Dobbins, Daniel Chubb, Ben Kinnersley, Malcolm G. Dunlop, Ian Tomlinson, Richard S. Houlston

PII: S0016-5085(16)35138-1
DOI: 10.1053/j.gastro.2016.09.041
Reference: YGAST 60735

To appear in: *Gastroenterology*
Accepted Date: 27 September 2016

Please cite this article as: Broderick P, Dobbins SE, Chubb D, Kinnersley B, Dunlop MG, Tomlinson I, Houlston RS, Validation of Recently Proposed Colorectal Cancer Susceptibility Gene Variants in an Analysis of Families and Patients—a Systematic Review, *Gastroenterology* (2016), doi: 10.1053/j.gastro.2016.09.041.

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], Richard S Houlston[1,4]

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.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

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)[1-7]. To search for independent evidence of a role in CRC risk we analyzed sequencing data on 863 familial CRC cases and 1,604 controls[8]. All cases were diagnosed aged ≤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)[6]. 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[9]. 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)[3]. No disruptive *RPS20* variants have been catalogued by the Exome-Aggregation-Consortium (ExAC), which contains WES data for 60,706 individuals of diverse ancestries[10] 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)[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 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[11]; 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[4], 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*[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&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 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[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 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

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.

		Disruptive mutations (stop-gain, frameshift)			Damaging mutations (disruptive, predicted-damaging, splice acceptor/donors			All coding non-synonymous variants		
Gene	**Previously Reported**	**Cases**	**Control**	P_{Fisher}	**Cases**	**Control**	P_{Fisher}	**Cases**	**Control**	P_{Fisher}
BUB1	Disruptive	0	4	0.31	1	8	0.17	18	30	0.76
BUB3	Missense	0	2	0.55	0	4	0.31	1	5	0.67
FAN1	Disruptive /Missense	0	2	0.55	15	17	0.19	32	45[#]	0.23
FANCM	Disruptive /Missense	5	1	**0.02**	23	33	0.33	51[$]	67[$]	0.06
LRP6 (BPD*)	Missense	0	0	-	6 (4)	17 (13)	0.51 (0.45)	17 (8)	37 (21)	0.67
PTPN12	Missense	0	1	1.00	6	5	0.21	12	9	**0.04**
RPS20	Disruptive	1	0	0.35	2	0	0.12	2	0	0.12
TP53	Missense	1	0	0.35	1	1	1.00	1	4	0.66

* Number of variants within β-Propellor domain. All 3 variants identified by de Voer *et al* were within BPD.

Total number of variants in controls = 46; 1 sample has 2 *FAN1* missense

$ Totals number of variants in cases = 52, in controls =69; 3 samples have 2 *FANCM* missense

SUPPLEMENTARY METHODS AND MATERIALS

METHODS

INDEPENDENT EVALUATION

Whole-exome sequencing data: To search for independent evidence and to contextualize the impact of each purported CRC gene we made use of recently published whole-exome sequencing (WES) data on 1,006 early-onset familial CRC cases and 1,609 healthy controls[1]. Cases were of European Ancestry recruited to the UK National Study of Colorectal Cancer Genetics (NSCCG)[2]. All cases were diagnosed with CRC aged ≤55 and had at least one first-degree relative diagnosed with CRC. Controls were individuals with no history of malignancy selected from the 1958 Birth Cohort (1958BC), a longitudinal study following the lives of people born in England, Scotland and Wales during the week of 3-9 March 1958[3]. Full details of WES have been published previously. Briefly, paired end fastq files were aligned to build 37 (hg19) of the human reference genome and alignments were processed using the Genome Analysis Tool Kit (GATKv3) pipeline according to best practices[4]. The Variant Effect Predictor[5] was used to provide annotations on the predicted impact of each variant together with functional classifications and assessment of deleteriousness from the CONDEL[6] algorithm. Samples (cases and controls) with a variant in an established high-penetrance CRC gene which was predicted to be disruptive (stop-gain, frameshift) or previously catalogued as being pathogenic or likely-pathogenic by InSiGHT (The International Society for Gastrointestinal Hereditary Tumours) were removed. Specifically: *APC:* 19 cases, 1 control; *MLH1:* 46 cases; *MSH2*: 46 cases; *MSH6*: 13 cases, 1 control; *MUTYH*: 9 cases; *PMS2*: 6 cases, 2 controls; *POLD1*: 1 case, 1 control; *POLE:* 3 cases; *BMPR1A, SMAD4*, *STK11:* 0 cases. Thus for the analysis presented in this manuscript we made use of whole-exome sequencing data on 863 cases and 1,604 controls.

Gene Burden Analysis: With currently attainable sample sizes, a studies power to detect a statistically significant association with any rare variant is typically weak. Here we use WES data described above, to look for an enrichment of variation in cases versus controls, for each postulated CRC gene as a whole. As the power of such comparison depends critically on the ability to distinguish between pathogenic and non-pathogenic variation, we defined and compared a number of variant classes: (1) Disruptive mutations (stop-gain, frameshift); (2) Disruptive and predicted damaging mutations (stop-gain, frameshift, missense predicted to be damaging by CONDEL, splice site acceptor/donors); (3) All coding non-synonymous variants. We assessed rare (minor allele frequency [MAF] <1%) and very rare (MAF<0.1%) mutations in each variant class. Comparisons were made using (a) all 863 cases (b) 159 cases with Amsterdam-II positive family histories (Amsterdam-I n=146). Thus in total we performed 12 comparisons for each gene.

Further analysis of the recurrent *FANCM* p.Arg1931Ter (rs144567652) variant: We studied the association of the recurrent variant *FANCM* p.Arg1931Ter (rs144567652) with CRC by analyzing published Illumina Infinium Human Exome BeadChip 12v1.0 or 12v1.1 exon array data[7]. Specifically, we made use of UK case/control data (excluding samples also included in our WES data) comprising: (i) 3,537 English CRC cases and 4,811 control patients; (ii) 2,015 Scottish CRC cases and 1,981 Scottish controls.

EVIDENCE FRAMEWORK

To assess the validity of purported CRC genes, accounting for varying study design, we collated the following evidence where appropriate (Supplementary Table 1):

(1) Where gene/variants were identified through the analysis of multiple members of a single family we evaluated the strength of segregation data – co-inheritance of the mutation with affection status (CRC or polyps) in the family. If not formally quantified in the published report we calculated non-parametric linkage (NPL) statistic *P*-values[8] using the family information provided.

(2) Where a specific CRC risk variant was reported: we looked for reported evidence of a statistically significant enrichment in CRC cases versus controls. In conjunction with

our WES data we examined frequency data on the mutation in an ethnically appropriate subset of the Exome-Aggregation-Consortium (ExAC) database[9]; a catalog of exome sequencing data for 60,706 individuals of diverse ancestries (non-Finnish European (NFE) 33,370 exomes, East Asian (EAS) 4,327 exomes, Finnish (FIN) 3,307 exomes).

(3) Where numerous variants are identified in a specific gene: we looked for evidence of gene burden testing in cases versus controls, and if performed, evidence of statistically significant enrichment of mutation in cases.

(4) Where recessive inheritance was suspected or indicated: homozygosity or compound-heterozgosity for pathogenic mutations in the proposed CRC gene was assessed in cases and controls.

(5) Computational data on the presumptive effect of the variant.

(6) Functional data – demonstration that mutation has a functional effect and the relevance to CRC biology.

(7) Other information – evidence of a highly-specific phenotype associated with CRC, evidence of somatic mutation of the wild-type allele in cancers from carriers consistent with tumor suppressor gene function.

REVIEW OF EXISTING LITERATURE FOR EACH GENE ASSESSED

NTHL1

Evidence for variation in *NTHL1* as a cause of recessive CRC has been provided by three unrelated Dutch families homozygous for the rare (ExAC NFE MAF=0.0023, homozygosity ~1/75,000) inactivating p.Gln90Ter mutation[10]. Multiple colorectal adenomas with or without CRC were diagnosed in all seven homozygotes. The tumor mutation spectrum was significantly enriched for C:G>T:A transitions, consistent with the mutation spectra observed in *NTHL1* double-knockout mice. Subsequent to this report, compound heterozygosity for inactivating *NTHL1* p.Gln90Ter/c.709+1G>A mutations was identified in a 41-year old Canadian woman diagnosed with polyposis, CRC and multiple tumors[11]. Tumors were again enriched for somatic C:G>T:A transitions.

Evidence Summary: Multiple reports associating homozygosity and compound-heterozygosity with CRC. Evidence of functional effect in CRC.

RPS20

In seven affected members (average age 52, range 24-75) of a four-generation Finnish Amsterdam-positive FCCTX family, Nieminen *et al.* identified a heterozygous 1-bp duplication, resulting in a frameshift and premature termination (p.Val50SerfsTer23), in *RPS20*[12]. The mutation, identified through genetic linkage analysis and WES, showed full co-segregation with microsatellite-stable early-onset CRC thus providing statistically significant evidence (reported LOD score=3.0; calculated NPL=5.35, *P*=0.0078) for germline mutation in *RPS20* as a cause of CRC. The mutation was absent in 292 population controls and is not reported in the ExAC database, which includes 3,307 Finnish individuals. Tumors from mutation carriers did not show loss of the wildtype allele (LOH^{WT}). Lymphoblastoid cells (LCLs) from cases carrying p.Val50SerfsTer23 mutation showed a marked increase in 21S pre-rRNAs compared to controls (*P*-value not calculated), consistent with a late pre-rRNA processing defect and suggestive of *RPS20* haploinsufficiency. Germline *RPS20* mutations were not found in 25 additional Finnish FCCTX families, 292 population controls or in tumor DNA from 50 primary CRC and 11 CRC cell lines.

Evidence Summary: Statistically significant evidence of segregation, absence of gene mutation in controls. Functional evidence in non-colon tissue. No evidence of somatic mutation or functional effect in CRC.

FANCM

By searching within tumorigenesis genes for rare/novel truncating mutations, which also showed LOH^{WT} within the tumor, Smith *et al.* identified the *FANCM* mutation p.Arg1931Ter (rs144567652) in one of 50 sporadic UK CRC cases[13]. As *FANCM* is functionally linked to *MSH2/MSH6* they sought further evidence for the role of this variant by genotyping an additional case-control series identifying the mutation in 1 of 2,207 CRC cases and 1 of 2,176 controls. The tumor of the additional case again showed LOH^{WT}. Combining discovery and replication samples showed no significant enrichment of the mutation in CRC (cases

2/2,257, controls 1/2,176, *P*=0.57). Smith *et al.* presented no segregation or functional data. However, a subsequent report by Peterlongo *et al.* proposing p.Arg1931Ter as a familial breast cancer risk factor showed that p.Arg1931Ter induces exon skipping resulting in decreased DNA repair activity in mouse embryonic fibroblast cells[14].

Evidence Summary: Loss of wild-type allele in tumors. Functional evidence in non-colon tissue. No evidence of segregation or functional effect in CRC. No significant enrichment of the mutation in CRC.

FAN1

Through WES of three individuals from a Spanish MMR-proficient, Amsterdam-positive CRC family Sequi *et al.* identified a novel *FAN1* truncating mutation p.Cys47Ter[15]. Evidence of segregation with CRC was limited (calculated NPL=0.95, *P*=0.25). Tumors developed by p.Cys47Ter mutation carrier showed no reduction in the expression of wild-type RNA or FAN1 protein. Screening an additional 247 Spanish Amsterdam/Bethesda positive cases for rare (MAF<0.01; dbSNP135) variants identified an additional truncating mutation (p.Arg952Ter) and three missense mutations (p.Asp140Thr and p.Arg591Trp - predicted to be damaging by SIFT and CONDEL algorithms; p.Pro340Ser - predicted to be benign). No *FAN1* mutations were identified in 250 population individuals without CRC. Whilst suggestive of an enrichment in cases for the overall burden of variation in *FAN1*, the size of the control population is insufficient to provide statistically robust support (combined 5/248 cases, 0/250 controls, *P*=0.061). Using all five families the calculated NPL segregation score was non-significant (NPL=1.05, *P*= 0.125, Supplementary Table 1). There was no evidence of *FAN1* LOH or somatic mutation in tumors from any of the five *FAN1* mutation carriers. LCLs derived from p.Cys47Ter and p.Asp140Thr carriers showed greater sensitivity to high doses of mitomycin C (MMC) compared to cells from a wild-type individual (p.Cys47Ter: *P*=0.01 Wilcoxon rank sum test; p.Asp140Thr: *P*-value not calculated). Transfection of a *FAN1* knockout human embryonic kidney cell line with p.Asp140Thr failed to reverse its MMC sensitivity.

Evidence Summary: Limited evidence of segregation and an increase in mutational burden in cases. Functional evidence for 2/4 mutations in non-colon tissue. No evidence of somatic mutation, LOH or functional effect in CRC.

TP53

Yurgelun *et al.* examined the frequency of rare germline *TP53* missense mutations in 457 patients with early-onset CRC (median age 36, range 15-40) and without a known hereditary cancer syndrome[16]. In six of the patients (1.3%), they identified missense changes in *TP53*. No comparison was made to the burden of TP53 mutations in controls.

Based on this data they concluded that the frequency of *TP53* mutations is comparable with the proportion of inherited CRC thought to be attributable to germline *APC* mutations. This is a false comparison:

(1) *TP53* missense variant are assumed to be deleterious. However, of the six variants they identified only one was predicted to be damaging by both SIFT and PolyPhen-2, with three being predicted benign by both algorithms. There was no other evidence that mutations had deleterious functional effect.

(2) The proportion of inherited CRC thought to be attributable to germline *APC* mutations (1%) is not the same as the frequency of samples with *APC* missense changes (in our original 1,006 cases, not screened for mutations in known genes, 94 [9.3%] had rare [MAF<1%] missense changes in *APC*).

Evidence Summary: No comparison was made to mutational burden in controls. No evidence (even *in silico*) mutations have functional effect. No evidence of segregation or somatic mutation in CRC.

BUB1* and *BUB3

BUB1: Disruptive mutations (p.Gln16Ter, p.Gln949Argfs) were identified in two of 23 Chinese early-onset (age at diagnosis ≤45) CRC cases; both variants were absent from 700 population controls but no comparison was made with the burden of mutations in controls[17]. No evidence of segregation was demonstrated and notably one of the mutation

carriers harbored a *MLH1* mutation (InSiGHT Class4: likely pathogenic), which unlike the *BUB1* variant was also carried by a sibling with polyps. The functional effect of mutations, LOH or somatic mutation within tumors was not assessed. No *BUB1* mutations were identified among 184 Netherlands/German cases.

Evidence Summary: Absence of identified mutations in controls, but no comparison made to mutational burden in controls. No functional evidence. No evidence of segregation or somatic mutation in CRC.

BUB3: WES identified a novel damaging (as predicted by PolyPhen-2 SIFT, HOPE) missense mutation (p.Phe264Leu) in one of 10 Dutch early-onset (age at diagnosis ≤45) CRC cases (0/23 Chinese cases)[17]**.** Sequencing *BUB3* in 174 Netherlands/German CRC cases identified two further missense variants (p.Lys21Asn, p.Arg149Gln) that were predicted (by at least one algorithm) to be damaging, although PolyPhen-2 and SIFT predicted p.Arg149Gln to be benign. All three variants were absent in 1,154 controls but no comparison was made with burden of mutations in controls. No evidence of segregation, LOH or somatic mutation. Lymphocytes and primary skin fibroblasts from p.Phe264Leu and p.Lys21Asn mutation carriers showed significant enrichment of aneuploidy and structural abnormalities versus controls (p.Arg149Gln was not assessed/presented).

Evidence Summary: Absence of identified mutations in controls, but no comparison made to mutational burden in controls. Functional evidence in non-colon tissue. No evidence of somatic mutation, LOH or functional effect in CRC.

LRP6* and *PTPN12

Using WES de Voer *et al.* looked for genes recurrently affected by damaging missense mutations, assessed using a single prediction algorithm PhyloP, in 55 Dutch non-polyposis MMR-proficient early-onset (age at diagnosis ≤45) CRC cases[18]:

LRP6: Damaging *LRP6* missense mutations (p.Trp239Leu, p.Asn789Ser, p.Thr867Ala) were identified in three cases. All three variants were within β-propeller domains. In mutation

carriers LRP6 protein showed no difference in protein expression or subcellular localization compared to wild-type. In Chinese Hamster Ovary cells p.Asn789Ser and p.Thr867Ala induced significant increases in WNT signaling activity (P<0.001). Analysis of 174 additional Netherlands/German CRC and 2,329 population controls identified no additional damaging missense mutations in cases, with 18 identified in controls including p.Thr867Ala in three controls. By using only the 55 original cases and the 2,329 controls, de Voer *et al.* reported significant increase in mutation burden in cases versus controls (cases 3/55, controls 18/2,329, P =0.01). However using all cases there was no significant enrichment of *LRP6* mutation in cases versus controls (cases 3/229, controls 18/2,329, P=0.43).

Evidence Summary: No increase in mutational burden in cases. Functional evidence in non-colon tissue. No evidence of somatic mutation, LOH or functional effect in CRC.

PTPN12: WES identified two damaging missense mutations (p.Arg522Met, p.Ser684Leu) in three of the 55 cases. Analysis of 174 additional Netherlands/German CRC and 2,329 population controls identified a new variant (p.Ala105Val) in one case and 11 variants in controls including previously identified p.Arg522Met in two controls and p.Ser684Leu in three controls. The burden of mutation is cases was significantly enriched versus controls (cases 4/229, controls 11/2329, P=0.039) albeit non-significant after adjustment for multiple testing (three candidate genes, P=0.12).

Evidence Summary: Limited evidence of an increase in mutational burden in cases. Variants not absent from controls. No evidence of somatic mutation, LOH or functional effect in CRC.

SUPPLEMENTARY FIGURES AND TABLES

Supplementary Figure 1: Forest plot of allelic odds ratio associated with *FANCM* p.Arg1931Ter (rs144567652) genotype and CRC. Studies [SMITH: original publication (2,207 cases, 2,176 controls)[13]; WES: whole-exome sequencing analyzed in this manuscript (863 cases, 1,604 controls); ENG: English Illumina Exome-BeadChip replication series (3,537 cases, 4,811 controls); SCOT: Scottish Illumina Exome-BeadChip replication series (2,015 cases, 1,981 controls)] were weighted according to the inverse of the variance of the log of the odds ratio (OR) calculated by unconditional logistic regression. Meta-analysis under a fixed-effects model was conducted using standard methods. Cochran's *Q* statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity were calculated. Horizontal lines indicate 95% confidence intervals (CIs). Boxes indicate OR point estimate; its area is proportional to the weight of the study. Diamond (and broken line) indicates overall summary estimate, with CI given by its width. Unbroken vertical line indicates null value (OR=1.0).

Supplementary Table 1: Evidence for genes and variants being associated with CRC risk. Analysis of the evidence presented in publications linking *NTHL1, RPS20, FANCM, FAN1, TP53, BUB1, BUB3, LRP6 and PTPN12* with the risk of developing CRC

Abbreviations: AOD: Age of CRC diagnosis; EAS: East Asian; FCCTX: Familial CRC Cancer Type X; FS: frameshift; Het: heterozygous; Hom: homozygous; LOH: loss of heterozygosity; MMC: Mitomycin C; MS: missense; MSS: lack of mismatch repair deficiency (MMR) tested through either microsatellite stability or no loss of MMR proteins; NFE: Non-Finnish European; NPL: non-parametric linkage; NT: Not tested; SG: stop-gain; TS: target sequencing; WES: whole exome sequencing; WT: wild-type

Supplementary Table 2: Gene Burden analysis. Number of WES cases (n=863) and controls (n=1,604) with rare mutations in genes suggested to increase CRC risk. We considered three sets of variants: Class-1, disruptive mutations -stop-gain, frameshift; Class-2, predicted damaging -disruptive plus missense predicted to be damaging by CONDEL and splice site acceptor/donors; Class-3, all coding non-synonymous variants. Tables show -A all cases -n=863 and controls -n=1,604 with very rare -MAF<0.1% mutations; -B Amsterdam-II positive cases -n=159 and controls with rare -MAF<1% mutations; -C Amsterdam-II positive cases and controls with very rare -MAF<0.1% mutations. For each gene and variant class, numbers of cases and controls were compared and *P*-values calculated using Fishers exact test.

Supplementary Table 3: ***BUB1, BUB3, FAN1, FANCM, LRP6, PTPN12, RPS20*** **and** ***TP53*** **variants -MAF<1% identified in 863 CRC cases and 1,604 controls.** -See excel file

REFERENCES

1. **Chubb D, Broderick P, Dobbins SE,** et al. Nat Commun 2016:7:11883.
2. Penegar S, et al. Br J Cancer 2007:97:1305-9.
3. Power C, et al. Int J Epidemiol 2006:35:34-41.
4. McKenna A, et al. Genome Res 2010:20:1297-303.
5. McLaren W, et al. Bioinformatics 2010:26:2069-70.
6. Gonzalez-Perez A et al. Am J Hum Genet 2011:88:440-9.
7. **Timofeeva MN, Kinnersley B**, et al. Sci Rep 2015:5:16286.
8. Kruglyak L et al. Am J Hum Genet 1996:58:1347-63.
9. Lek M, et al. Nature 2016:536:285-91.
10. Weren RD, et al. Nat Genet 2015:47:668-71.
11. Rivera B, et al. N Engl J Med 2015:373:1985-6.
12. Nieminen TT, et al. Gastroenterology 2014:147:595-598 e5.
13. Smith CG, et al. Hum Mutat 2013:34:1026-34.
14. Peterlongo P et al. Hum Mol Genet 2015:24:5345-55.
15. **Segui N, Mina LB,** et al. Gastroenterology 2015:149:563-6.
16. Yurgelun MB et al. JAMA Oncol 2015:1:214-21.
17. **de Voer RM, Geurts van Kessel A, Weren RD, Ligtenberg MJ,** et al. Gastroenterology 2013:145:544-7.
18. **de Voer RM, Hahn MM**, et al. PLoS Genet 2016:12:e1005880

Author names in bold designate shared co-first authorship

Supplementary Table 1: Evidence for genes and variants being associated with CRC risk. Analysis of the evidence presented in publications linking *NTHL1, RPS20, FANCM, FAN1, TP53, BUB1, BUB3, LRP6 and PTPN12* with the risk of developing CRC

Abbreviations: AOD: Age of CRC diagnosis; EAS: East Asian; FCCTX: Familial CRC Cancer Type X; FS: frameshift; Het: heterozygous; Hom: homozygous; LOH: loss of heterozygosity; MMC: Mitomycin C; MS: missense; MSS: lack of mismatch repair deficiency (MMR) tested through either microsatellite stability or no loss of MMR proteins; NFE: Non-Finnish European; NPL: non-parametric linkage; NT: Not tested; SG: stop-gain; TS: target sequencing; WES: whole exome sequencing; WT: wild-type

Gene	CRC Gene Burden	Control Gene Burden	Gene Functional Data	Chr	Position (GRCh37)	c.DNA Change	Protein Change	Class	dbSNP	Segregation	Variant Case Frequency	Variant Control Frequency	ExAC Allele Frequency	Other Information
NTHL1	WES Hom 3/51 APC-MUTYH mutation-negative polyposis patients	17 Het/ 2329		16	2096239	c.268C>T	p.Gln90Ter	SG	rs150766139	Family1:2/2 Family2:3/3 Family3:2/2		17 Het/ 2329	NFE: 0.0023	Tumors significantly enriched for C:G>T:A transitions
RPS20	WES 4 individuals Finnish FCCTX family; TS 0/25 Finnish FCCTX	NT	RPS20 depletion in HeLA cells and LCLs from patients carrying c.147dupA showed increase in 21S pre-rRNA vs controls	8	56986283	c.147dupA	Val50Serfs Ter23	FS		NPL=5.35, *P* =0.0078	1/26 FCCTX Family	0/584	Absent	No LOH (0/2)
FANCM	WES 1/50	NT		14	45667921	c.5791C>T	p.Arg1931Ter	SG	rs144567652	NT	2/2,258	1/2176	NFE: 0.0009	LOH WT 2/2
FAN1	WES 3 individuals Spanish FCCTX family. TS 4/247 Spanish FCCTX cases	0/250	LCL from p.C47Ter and p.Asp140Tyr carriers showed greater sensitivity to high doses of MMC. Transfection of HEK293T (Human Embryonic Kidney) FAN1 with p.Asp140Tyr failed to reverse its MMC sensitivity	15	31197007	c.141C>A	p.Cys47Ter	SG	rs144469584	NPL=0.94, *P* =0.25	1/248 FCXX Family	0/538	Absent	No LOH/somatic mutation. No reduction in FAN1 RNA or protein.
				15	31197284	c.418G>T	p.Asp140Tyr	MS	rs761776412	P/C, NPL=0, *P* =1	1/248 FCXX Family	0/250	NFE: 1.50E-5	Predicted to be damaging by SIFT/CONDEL. No LOH/somatic mutation.
				15	31197884	c.1018C>T	p.Pro340Ser	MS	rs771206220	NPL=1.05, *P* =0.125	1/248 FCXX Family	0/250	NFE: 1.50E-5	Predicted to be benign by SIFT/CONDEL

Gene	CRC Gene Burden	Control Gene Burden	Gene Functional Data	Chr	Position (GRCh37)	c.DNA Change	Protein Change	Class	dbSNP	Segregation	Variant Case Frequency	Variant Control Frequency	ExAC Allele Frequency	Other Information
FAN1 (cont'd)				15	31206254	c.1771C>T	p.Arg591Trp	MS	rs377418523	NT	1/248 FCXX Family	0/250	NFE: 1.50E-5	Predicted to be damaging by PolyPhen-2/SIFT/CONDEL
				15	31222812	c.2854C>T	p.Arg952Ter	SG	rs184745027	NPL=0.70, *P* =0.25	1/248 FCXX Family	0/250	NFE: 9.02E-5	
TP53	TS 6/457 North-American, Australian, New Zealand non-polyposis, AOD≤40	NT		17	7572973	c.1136G>A	p.Arg379His	MS		NT	1/457	NT	Absent	Predicted to be benign by PolyPhen-2/SIFT
				17	7577069	c.869G>A	p.Arg290His	MS	rs55819519	NT	1/457	NT	NFE: 0.0002	Predicted to be benign by PolyPhen-2/SIFT
				17	7577088	c.850A>T	p.Thr284Ser	MS	s144340710	NT	1/457	NT	Absent	Predicted to be possibly damaging by PolyPhen-2, benign by SIFT
				17	7577091	c.847C>T	p.Arg283Cys	MS	rs149633775	NT	1/457	NT	NFE: 0.0002	Predicted to be benign by PolyPhen-2, possibly damaging by SIFT
				17	7577577	c.704A>G	p.Asn235Ser	MS	rs144340710	NT	1/457	NT	NFE: 0.0003	Predicted to be benign by PolyPhen-2/SIFT
				17	7578475	c.455C>T	p.Pro152Leu	MS	rs587782705	NT	1/457	NT	NFE: 4.50E-5	Predicted to be damaging by PolyPhen-2/SIFT
BUB1	WES 2/23 Han Chinese ≤40; WES 0/10 Dutch non-polyposis MSS ≤40; TS 0/174 non-polyposis MSS Dutch/German CRC	NT	Disruption of *BUB1* exon 1 in HCT- 116 (MSI human CRC) caused chromosomal segregation defects	2	111398721	c.2844delC	pGln949Argfs	FS		NT	1/23 Chinese; 0/184 European	0/700	Absent	
				2	111431923	c.46C>T	p.Gln16Ter	SG		No	1/23 Chinese; 0/184 European	0/700	Absent	Carries MLH1 splice donor mutation which segregates with colorectal tumors

Gene	CRC Gene Burden	Control Gene Burden	Gene Functional Data	Chr	Position (GRCh37)	c.DNA Change	Protein Change	Class	dbSNP	Segregation	Variant Case Frequency	Variant Control Frequency	ExAC Allele Frequency	Other Information
BUB3	WES 0/23 Han Chinese ≤40; WES 1/10 Dutch non-polyposis MSS ≤40; TS 2/174 non-polyposis MSS Dutch/German CRC	NT		10	124914496	c.63G>C	pLys21Asn	MS		NT	0/23 Chinese; 1/184 European	0/1154		Predicted to be damaging by PolyPhen-2/SIFT/HOPE
				10	124919951	c.446G>A	Arg149Gln	MS	rs371545161	NT	0/23 Chinese; 1/184 European	0/1154	NFE: 7.52E-5 EAS: Absent	Predicted to be damaging by HOPE, benign by PolyPhen-2/SIFT/HOPE
				10	124922163	c.790T>C	p.Phe264Leu	MS		NT	0/23 Chinese; 1/184 European	0/1154	Absent	Predicted to be damaging by PolyPhen-2/SIFT/HOPE
LRP6	WES 3/55 Dutch non-polyposis MSS AOD≤45; TS 0/174 non-polyposis MSS Dutch/German CRC	18/ 2,329	In Chinese Hamster Ovary cells: no effect on LRP6 protein expression or localization; overexpression of p.Asn789Ser and p.Thr867Ala induced increased WNT signaling vs WT	12	12311955	c.2599A>G	p.Thr867Ala	MS	rs141458215	NT	WES: 1/55; TS: 0/174	3/2329	NFE: 0.0002	Predicted to be damaging by PhyloP
				12	12312812	c.2366A>G	p.Asn789Ser	MS		NT	WES: 1/55; TS: 0/174	0/2329	Absent	Predicted to be damaging by PhyloP
				12	12339985	c.716G>T	p.Trp239Leu	MS		NT	WES: 1/55; TS: 0/174	0/2329	Absent	Predicted to be damaging by PhyloP
PTPN12	WES 3/55 Dutch non-polyposis MSS ≤45; TS 1/174 non-polyposis MSS Dutch/German CRC	11/ 2,329		7	77212900	c.314C>T	p.Ala105Val	MS		NT	WES: 0/55; TS: 1/174	0/2329	Absent	Predicted to be damaging by PhyloP
				7	77256561	c.1565G>T	p.Arg522Met	MS	rs537562368	NT	WES: 1/55; TS: 0/174	2/2329	NFE: 1.50E-5	Predicted to be damaging by PhyloP
				7	77261719	c.2051C>T	p.Ser684Leu	MS	rs201001953	NT	WES: 2/55; TS: 0/174	3/2329	NFE: 0.0012	Predicted to be damaging by PhyloP

Supplementary Table 2: Gene Burden analysis. Number of WES cases (n=863) and controls (n=1,604) with rare mutations in genes suggested to increase CRC risk. We considered three sets of variants: Class-1, disruptive mutations -stop-gain, frameshift; Class-2, predicted damaging -disruptive plus missense predicted to be damaging by CONDEL and splice site acceptor/donors; Class-3, all coding non-synonymous variants. Tables show -A all cases -n=863 and controls -n=1,604 with very rare -MAF<0.1% mutations; -B Amsterdam-II positive cases -n=159 and controls with rare -MAF<1% mutations; -C Amsterdam-II positive cases and controls with very rare -MAF<0.1% mutations. For each gene and variant class, numbers of cases and controls were compared and *P*-values calculated using Fishers exact test.

A

	Class-1			Class-2			Class-3		
Gene	**Cases**	**Controls**	P_{Fisher}	**Cases**	**Controls**	P_{Fisher}	**Cases**	**Controls**	P_{Fisher}
BUB1	0	4	0.31	1	8	0.17	5	16	0.36
BUB3	0	2	0.55	0	4	0.31	1	5	0.67
FAN1	0	2	0.55	3	5	1.00	7	10[#]	0.62
FANCM	1	0	0.35	7	8	0.42	14[$]	21	0.59
*LRP6 -BPD**	0	0	-	3 -1	9 -6	0.56	8 -2	20 -11	0.55
PTPN12	0	1	1.00	3	3	0.43	8	7	0.17
RPS20	1	0	0.35	2	0	0.12	2	0	0.12
TP53	1	0	0.35	1	1	1.00	1	3	1.00

*Number of variants within β-Propellor domain
Total number of variants in controls = 11; 1 sample has 2 *FAN1* missense
$ Total number of variants in cases = 15; 1 sample has 2 *FANCM* missense

B

	Class-1			Class-2			Class-3		
Gene	**Cases**	**Controls**	P_{Fisher}	**Case**	**Control**	P_{Fisher}	**Case**	**Control**	P_{Fisher}
BUB1	0	4	1.00	1	8	0.58	4	30	0.54
BUB3	0	2	1.00	0	4	1.00	0	5	1.00
FAN1	0	2	1.00	2	17	0.69	6	45	0.46
FANCM	0	1	1.00	1	33	0.36	7[$]	67[$]	0.84
*LRP6 -BPD**	0	0	-	1 -1	17 -13	1.00	2 -2	37 -21	0.57
PTPN12	0	1	1.00	1	5	0.44	3	9	0.09
RPS20	0	0	-	1	0	0.09	1	0	0.09
TP53	0	0	-	0	1	1.00	0	4	1.00

*Number of variants within β-Propellor domain
$ Total number of variants in cases = 8 in controls = 69; 3 samples have 2 *FANCM* missense

C

	Class-1			Class-2			Class-3		
Gene	**Cases**	**Controls**	P_{Fisher}	**Cases**	**Controls**	P_{Fisher}	**Cases**	**Controls**	P_{Fisher}
BUB1	0	4	1.00	1	8	0.58	1	16	1.00
BUB3	0	2	1.00	0	4	1.00	0	5	1.00
FAN1	0	2	1.00	0	5	1.00	2	10	0.30
FANCM	0	0	-	0	8	1.00	2[$]	21	1.00
*LRP6 -BPD**	0	0	-	0	9 -6	1.00	0	20 -11	0.25
PTPN12	0	1	1.00	0	3	1.00	1	7	0.53
RPS20	0	0	-	1	0	0.09	1	0	0.09
TP53	0	0	-	0	1	1.00	0	3	1.00

*Number of variants within β-Propellor domain

$ Total number of variants in cases = 3; 1 sample has 2 *FANCM* missense

Gene	Chr	Position (GRCh37)	Ref	Alt	CONSEQUENCE	Class	c.DNA Change	Protein Change	Control_MAF	Case_MAF	ALT_Control	ALT_Cases	ALT_Ams	ExAC_Freq	ExAC_NFE	1000G_ALL	1000G_EUR	Condel	CADD_PHRED	SIFT_score	SIFT_pred	Polyphen2_HDIV_score	Polyphen2_HDIV_pred	dbSNP	COSMIC_ID	COSMIC_DIS	ClinVar_SIG	ClinVar_DIS
TP53	17	7573948	C	G	missense_variant	3	c.1079G>C	p.Gly360Ala	0.0004	0	2	0		0.0003	0.0001	0.0002	0.001	neutral(0.002)	0.014	0.4	T	0	B	rs35993958	.	.	unknown	Hereditary_cancer-predisposing_syndrome;Hereditary_cancer-predisposing_syndrome
TP53	17	7573964	C	T	missense_variant	3	c.1063G>A	p.Ala355Thr	0.0004	0	1	0		.	.	.	.	neutral(0.064)	13.36	0.04	D	0.003	B	.	.	.	.	.
TP53	17	7577106	G	A	missense_variant	2	c.832C>T	p.Pro278Ser	0.0003	0	1	0		.	.	.	.	deleterious(0.765)	31	0	D	1	D	.	COSM1646810;CO	3(ovary);6(large_intestin	.	.
TP53	17	7579485	C	A	stop_gained	1	c.202G>T	p.Glu68Ter	0	0.0005	0	1		.	.	.	.	0	15.2	0.78	T	.	.	.	COSM1646879;CO	5(lung);1(pancreas);5(livi	.	.
LRP6	12	12274102	G	C	missense_variant	3	c.4800C>G	p.His1600Gln	0	0.0005	0	1		1.67E-05	.	.	.	neutral(0.034)	13.25	0.03	D	0.997	D	rs138902458	.	.	.	.
LRP6	12	12277562	C	T	missense_variant	2	c.4484G>A	p.Arg1495Gln	0.0003	0	1	0		9.06E-05	0.0001	.	.	deleterious(0.682)	36	0.17	T	0.999	D	rs149558764	.	.	.	.
LRP6	12	12278243	G	A	missense_variant	3	c.4436C>T	p.Thr1479Ile	0.0016	0	5	0		0.0002	0.0003	.	.	neutral(0.029)	16.56	0.25	T	0.403	B	rs144175121	.	.	.	.
LRP6	12	12278309	C	A	missense_variant	2	c.4370G>T	p.Gly1457Val	0.0003	0	1	0		4.13E-05	7.51E-05	.	.	deleterious(0.482)	22.9	0.25	T	0.991	D	rs374072779	.	.	.	.
LRP6	12	12279642	A	G	missense_variant	3	c.4295T>C	p.Leu1432Ser	0.0007	0	2	0		1.65E-05	3.00E-05	.	.	neutral(0.013)	13.62	0.78	T	0.557	P	.	.	.	.	.
LRP6	12	12279723	T	A	missense_variant	3	c.4214A>T	p.Asp1405Val	0.0006	0.0005	2	1		3.30E-05	5.99E-05	.	.	neutral(0.034)	14.54	0.05	D	0.343	B	.	.	.	.	.
LRP6	12	12279793	C	A	missense_variant	3	c.4144G>T	p.Val1382Phe	0.0003	0.0017	1	3		0.0009	0.0011	0.0012	0.001	neutral(0.044)	15	0.21	T	0.04	B	rs139480047	.	.	.	.
LRP6	12	12283731	T	G	missense_variant	2	c.4067A>C	p.Asp1356Ala	0.0003	0	1	0		1.66E-05	3.02E-05	.	.	deleterious(0.945)	23.1	0	D	0.998	D	rs376168250	.	.	.	.
LRP6	12	12283783	T	G	missense_variant	3	c.4015A>C	p.Ile1339Leu	0.0003	0	1	0		.	.	.	.	neutral(0.279)	15.88	0.03	D	0.05	B	.	.	.	.	.
LRP6	12	12283804	G	A	missense_variant	2	c.3994C>T	p.Arg1332Cys	0	0.0005	0	1		6.62E-05	1.51E-05	.	.	deleterious(0.634)	22.4	0.02	D	0.992	D	.	.	.	.	.
LRP6	12	12284815	C	T	missense_variant	3	c.3910G>A	p.Ala1304Thr	0.0003	0	1	0		5.77E-05	5.99E-05	.	.	neutral(0.017)	10.9	0.58	T	0.053	B	rs376223132	.	.	.	.
LRP6	12	12284821	C	T	missense_variant	3	c.3904G>A	p.Asp1302Asn	0	0.0006	0	1		.	.	.	.	neutral(0.461)	32	0.32	T	0.799	P	.	.	.	.	.
LRP6	12	12284914	C	T	missense_variant	2	c.3811G>A	p.Asp1271Asn	0	0.0005	0	1		.	.	.	.	deleterious(0.865)	35	0.09	T	1	D	.	.	.	.	.
LRP6	12	12288226	G	C	missense_variant	2	c.3616C>G	p.Pro1206Ala	0.0003	0	1	0		4.12E-05	6.00E-05	.	.	deleterious(0.476)	25.1	0.08	T	0.996	D	rs373983720	.	.	.	.
LRP6	12	12291339	T	G	missense_variant	3	c.3527A>C	p.Glu1176Ala	0.0003	0.0011	1	2		0.0001	0.0002	.	.	neutral(0.005)	14.56	0.58	T	0.005	B	rs145672862	.	.	.	.
LRP6	12	12301951	T	A	missense_variant	2	c.3131A>T	p.Asp1044Val	0.0003	0	1	0		8.24E-06	1.50E-05	.	.	deleterious(0.543)	20.9	0.03	D	0.838	P	.	.	.	.	.
LRP6	12	12302006	G	A	missense_variant	2	c.3076C>T	p.Arg1026Cys	0.0013	0	4	0		0.0001	0.0001	0.0008	.	deleterious(0.869)	23.3	0	D	1	D	rs140679529	COSM180991	1(endometrium);1(large_	.	.
LRP6	12	12303837	T	C	missense_variant	3	c.2927A>G	p.Asp976Gly	0.0003	0	1	0		1.65E-05	3.00E-05	.	.	neutral(0.050)	17.25	0.38	T	0.396	B	rs368801975	.	.	.	.
LRP6	12	12303861	C	T	missense_variant	3	c.2903G>A	p.Arg968Gln	0.0003	0	1	0		8.24E-06	0	.	.	neutral(0.313)	36	0.63	T	0.999	D	.	.	.	.	.
LRP6	12	12311913	C	T	missense_variant	2	c.2641G>A	p.Val881Ile	0.0003	0	1	0		8.24E-06	1.50E-05	.	.	deleterious(0.790)	32	0.54	T	0.999	D	COSM1561792	1(central_nervous_syste	.	.	.
LRP6	12	12311955	T	C	missense_variant	2	c.2599A>G	p.Thr867Ala	0.0003	0.0006	1	1		0.0001	0.0002	.	.	deleterious(0.471)	18.01	0.37	T	0.857	P	rs141458215	.	.	.	.
LRP6	12	12315203	C	T	missense_variant	2	c.2203G>A	p.Asp735Asn	0.0003	0	1	0		0.0001	0.0002	0.0002	0.001	deleterious(0.756)	36	0.51	T	1	D	rs149258721	.	.	.	.
LRP6	12	12315310	A	C	missense_variant	2	c.2096T>G	p.Val699Gly	0.0003	0	1	0		.	.	.	.	deleterious(0.877)	25.4	0	D	1	D	.	.	.	.	.
LRP6	12	12317322	G	C	missense_variant	2	c.1937C>G	p.Ser646Cys	0.0003	0	1	0		.	.	.	.	deleterious(0.801)	19.86	0.04	D	1	D	.	.	.	.	.
LRP6	12	12317387	C	G	missense_variant	3	c.1872G>C	p.Met624Ile	0	0.0005	0	1		.	.	.	.	neutral(0.060)	18.58	0.17	T	0.08	B	.	.	.	.	.
LRP6	12	12318150	C	T	missense_variant	2	c.1625G>A	p.Gly542Asp	0	0.0005	0	1		8.24E-05	0.0001	.	.	deleterious(0.777)	26.1	0.3	T	0.999	D	rs139800650	.	.	.	.
LRP6	12	12332842	C	T	missense_variant	3	c.1447G>A	p.Val483Ile	0.0003	0.0005	1	1		0.0162	0.0002	0.053	0.001	neutral(0.245)	17.25	0.36	T	0.013	B	rs7975614	.	.	.	.
LRP6	12	12334037	T	C	missense_variant	3	c.1313A>G	p.Lys438Arg	0	0.0006	0	1		.	.	.	.	neutral(0.043)	13.7	0.34	T	0.091	B	.	.	.	.	.
LRP6	12	12334052	T	C	missense_variant	2	c.1298A>G	p.Asn433Ser	0	0.0005	0	1		2.47E-05	4.50E-05	.	.	deleterious(0.634)	19.52	0.07	T	0.982	D	rs397515473	.	.	pathogenic	Coronary_artery_disease\x2c_autosomal_dominant_2
LRP6	12	12334128	T	C	missense_variant	3	c.1222A>G	p.Ile408Val	0.0003	0	1	0		.	.	.	.	neutral(0.261)	14.16	0.14	T	0.876	P	.	.	.	.	.
LRP6	12	12334248	C	T	missense_variant	2	c.1102G>A	p.Asp368Asn	0.0003	0	1	0		8.24E-06	1.50E-05	.	.	deleterious(0.798)	27.6	0.01	D	0.996	D	.	.	.	.	.
LRP6	12	12334271	C	T	missense_variant	2	c.1079G>A	p.Arg360His	0.0003	0.0011	1	1		0.0001	0.0002	.	.	deleterious(0.667)	33	0.13	T	1	D	rs141212743	.	.	pathogenic	Coronary_artery_disease\x2c_autosomal_dominant_2
LRP6	12	12339925	C	T	missense_variant	3	c.776G>A	p.Arg259His	0.0003	0	1	0		4.12E-05	3.00E-05	.	.	neutral(0.383)	18.13	0.13	T	0.983	D	.	.	.	.	.
LRP6	12	12397266	A	T	missense_variant	3	c.379T>A	p.Ser127Thr	0.0003	0	1	0		0.002	1.50E-05	0.006	.	neutral(0.000)	2.447	0.73	T	0	B	rs17848270	.	.	.	.
LRP6	12	12397530	C	T	missense_variant	3	c.115G>A	p.Gly39Ser	0.0003	0	1	0		8.24E-05	0.0001	.	.	neutral(0.002)	10.42	0.74	T	0.099	B	rs147654774	.	.	.	.
LRP6	12	12397550	C	T	missense_variant	2	c.95G>A	p.Arg32Gln	0.0003	0	1	0		.	.	.	.	deleterious(0.740)	25.2	0.07	T	0.999	D	.	COSM936449	1(endometrium)	.	.
FAN1	15	31196876	G	A	missense_variant	3	c.10G>A	p.Glu4Lys	0	0.0005	0	1		8.38E-06	1.53E-05	.	.	neutral(0.064)	18.35	0.05	D	0.775	P	rs369398471	.	.	.	.
FAN1	15	31196946	C	A	missense_variant	3	c.80C>A	p.Ser27Tyr	0.0003	0	1	0		.	.	.	.	neutral(0.406)	14.26	0.05	D	0.01	B	.	.	.	.	.
FAN1	15	31196961	C	T	missense_variant	3	c.95C>T	p.Ser32Leu	0.0006	0	2	0		3.30E-05	6.00E-05	.	.	neutral(0.256)	17.74	0.01	D	0.76	P	.	.	.	.	.
FAN1	15	31197015	T	G	missense_variant	2	c.149T>G	p.Met50Arg	0.0031	0.0061	10	10		0.002	0.0026	0.0018	0.006	deleterious(0.757)	22.8	0.01	D	0.999	D	rs148404807	.	.	.	.
FAN1	15	31197266	G	A	missense_variant	3	c.400G>A	p.Val134Met	0.0003	0	1	0		4.13E-05	7.51E-05	0.0002	0.001	neutral(0.046)	9.959	0.27	T	0.471	P	rs144046046	.	.	.	.
FAN1	15	31197291	T	C	missense_variant	3	c.425T>C	p.Leu142Pro	0	0.0005	0	1		.	.	.	.	neutral(0.035)	10.34	0.26	T	0.465	P	.	.	.	.	.
FAN1	15	31197300	G	A	missense_variant	3	c.434G>A	p.Arg145His	0.0019	0.002	6	3		0.0012	0.0018	0.0002	0.001	neutral(0.027)	10.99	0.31	T	0.025	B	rs146408181	.	.	.	.
FAN1	15	31197584	G	A	missense_variant	3	c.718G>A	p.Glu240Lys	0.0047	0.0065	15	9		0.0035	0.0052	0.0012	0.003	neutral(0.042)	11	0.27	T	0.154	B	rs150748572	.	.	.	.
FAN1	15	31197623	C	G	missense_variant	3	c.757C>G	p.Leu253Val	0	0.0005	0	1		.	.	.	.	neutral(0.032)	13.21	0.92	T	0.986	D	.	.	.	.	.
FAN1	15	31200356	C	T	missense_variant	2	c.1270C>T	p.Arg424Cys	0.0003	0	1	0		1.65E-05	3.00E-05	.	.	deleterious(0.638)	17.04	0	D	0.995	D	rs368417113	.	.	.	.
FAN1	15	31200396	A	G	missense_variant	3	c.1310A>G	p.Glu437Gly	0.0006	0.0005	2	1		0.0041	0.0007	0.012	0.001	neutral(0.013)	7.126	0.35	T	0.008	B	rs79759675	.	.	.	.
FAN1	15	31206123	G	A	missense_variant	2	c.1640G>A	p.Arg547His	0.0003	0	1	0		8.25E-06	1.50E-05	.	.	deleterious(0.892)	33	0.04	D	1	D	.	.	.	.	.
FAN1	15	31210489	C	G	missense_variant	3	c.1934C>G	p.Pro645Arg	0	0.0007	0	1		.	.	.	.	neutral(0.421)	17.92	0.3	T	0.991	D	.	.	.	.	.
FAN1	15	31212765	C	T	missense_variant	2	c.1961C>T	p.Pro654Leu	0.0006	0.0011	2	2		0.0002	0.0004	.	.	deleterious(0.906)	25.1	0.03	D	1	D	rs144081053	.	.	.	.
FAN1	15	31214480	T	G	missense_variant	2	c.2095T>G	p.Tyr699Asp	0	0.0012	0	2		.	.	.	.	deleterious(0.861)	22.7	0.02	D	1	D	.	.	.	.	.
FAN1	15	31214513	C	T	stop_gained	1	c.2128C>T	p.Arg710Ter	0.0007	0	2	0		2.47E-05	1.50E-05	.	.	0	42	1	T	.	.	rs199845994	.	.	.	.
FAN1	15	31214514	G	A	missense_variant	2	c.2129G>A	p.Arg710Gln	0.0003	0	1	0		1.65E-05	.	.	.	deleterious(0.843)	35	0	D	1	D	.	.	.	.	.
FAN1	15	31221509	G	A	missense_variant	3	c.2696G>A	p.Arg899Gln	0.0004	0.0006	1	0		2.47E-05	4.50E-05	.	.	neutral(0.010)	2.025	0.64	T	0.273	B	.	.	.	.	.
FAN1	15	31221566	G	A	missense_variant	3	c.2753G>A	p.Ser918Asn	0.0004	0.0006	1	0		1.65E-05	3.00E-05	.	.	neutral(0.038)	12.69	1	T	0.107	B	.	.	.	.	.
FAN1	15	31222830	C	T	missense_variant	2	c.2872C>T	p.Leu958Phe	0	0.0006	0	1		8.27E-06	1.50E-05	0.0002	0.001	deleterious(0.506)	20.5	0.15	T	0.997	D	rs147984615	.	.	.	.
FANCM	14	45605302	C	A	missense_variant	3	c.68C>A	p.Pro23Gln	0	0.0005	0	1		1.69E-05	3.09E-05	.	.	neutral(0.043)	15.42	0.2	T	0.929	P	rs377031191	.	.	.	.
FANCM	14	45605336	C	A	missense_variant	3	c.102C>A	p.Ser34Arg	0.0003	0	1	0		.	.	.	.	neutral(0.174)	20.5	0.08	T	0.838	P	.	.	.	.	.
FANCM	14	45605397	G	A	missense_variant	2	c.163G>A	p.Asp55Asn	0.0006	0.0016	2	2		0.0004	0.0008	.	.	deleterious(0.649)	35	0.02	D	0.993	D	rs148017562	COSM3718474	1(upper_aerodigestive_t	.	.
FANCM	14	45605405	G	C	missense_variant	3	c.171G>C	p.Leu57Phe	0.0029	0.0064	9	10		0.0016	0.0025	0.0004	0.002	neutral(0.414)	15.37	0.16	T	0.078	B	rs142007602	.	.	.	.
FANCM	14	45605503	C	T	missense_variant	2	c.269C>T	p.Pro90Leu	0.0013	0	4	0		0.0002	0.0004	.	.	deleterious(0.757)	29.1	0	D	0.916	P	rs142904668	.	.	.	.
FANCM	14	45606290	C	T	missense_variant	3	c.527C>T	p.Thr176Ile	0.0032	0.0041	10	7		0.0044	0.0056	0.0014	0.005	neutral(0.050)	10.31	0.22	T	0.003	B	rs77374493	.	.	.	.
FANCM	14	45618149	T	C	missense_variant	3	c.869T>C	p.Ile290Thr	0	0.0005	0	1		4.97E-05	9.03E-05	.	.	neutral(0.449)	15.03	0.04	D	0.147	B	rs377303950	.	.	.	.
FANCM	14	45618154	C	A	missense_variant	2	c.874C>A	p.Pro292Thr	0	0.0006	0	1		.	.	.	.	deleterious(0.747)	18.46	0.01	D	0.896	P	.	.	.	.	.
FANCM	14	45618161	G	A	missense_variant	2	c.881G>A	p.Gly294Asp	0.0003	0	1	0		.	.	.	.	deleterious(0.509)	18.3	0.02	D	0.212	B	.	.	.	.	.
FANCM	14	45623953	T	C	missense_variant	2	c.1237T>C	p.Tyr413His	0.0023	0.0005	7	1		0.0008	0.0012	0.0002	.	deleterious(0.789)	18.7	0.05	D	0.966	D	rs138225703	.	.	.	.
FANCM	14	45624629	G	A	missense_variant	2	c.1363G>A	p.Glu455Lys	0	0.0005	0	1		.	.	.	.	deleterious(0.698)	28	0.07	T	0.753	P	.	.	.	.	.
FANCM	14	45628364	A	G	missense_variant	2	c.1462A>G	p.Ser488Gly	0.0006	0	2	0		2.56E-05	4.69E-05	.	.	deleterious(0.902)	24	0.01	D	0.819	P	rs374988842	.	.	.	.
FANCM	14	45628452	C	T	missense_variant	3	c.1550C>T	p.Thr517Met	0.0003	0.0006	1	1		3.33E-05	4.55E-05	.	.	neutral(0.024)	10.89	0.29	T	0.165	B	.	.	.	.	.
FANCM	14	45633577	C	T	missense_variant	2	c.1597C>T	p.Arg533Cys	0	0.0006	0	1		0.0004	0.0006	.	.	deleterious(0.592)	27.9	0	D	0.991	D	rs146151355	COSM1721918	2(NS)	.	.
FANCM	14	45633616	G	A	missense_variant	2	c.1636G>A	p.Gly546Ser	0	0.0005	0	1		4.95E-05	7.50E-05	.	.	deleterious(0.935)	36	0.01	D	1	D	.	.	.	.	.
FANCM	14	45636337	G	A	missense_variant	3	c.1973G>A	p.Arg658Gln	0.0003	0	1	0		8.28E-06	1.50E-05	.	.	neutral(0.297)	16.19	0.3	T	0.944	P	.	.	.	.	.
FANCM	14	45639927	T	C	missense_variant	2	c.2138T>C	p.Leu713Ser	0.0004	0	1	0		.	.	.	.	deleterious(0.794)	11.11	0.02	D	0.985	D	.	.	.	.	.
FANCM	14	45642333	A	G	missense_variant	2	c.2236A>G	p.Thr746Ala	0	0.0005	0	1		8.24E-06	1.50E-05	.	.	deleterious(0.585)	14.04	0.09	T	0.819	P	.	.	.	.	.
FANCM	14	45642358	G	A	missense_variant	2	c.2261G>A	p.Arg754Gln	0.0003	0	1	0		1.65E-05	1.50E-05	0.0002	.	deleterious(0.884)	33	0.04	D	1	D	rs529256568	.	.	.	.
FANCM	14	45644346	G	A	missense_variant	3	c.2389C>G	p.Pro797Ala	0.0003	0.0005	1	1		8.26E-06	1.50E-05	.	.	neutral(0.002)	0.056	0.7	T	0.031	B	rs144771929	.	.	.	.
FANCM	14	45644458	A	G	missense_variant	2	c.2501A>G	p.Glu834Gly	0.0006	0	2	0		.	.	.	.	deleterious(0.715)	14.72	0.01	D	0.979	D	.	.	.	.	.
FANCM	14	45644709	A	G	missense_variant	3	c.2752A>G	p.Lys918Glu	0.0003	0	1	0		2.48E-05	4.51E-05	.	.	neutral(0.000)	0	0.86	T	0	B	.	.	.	.	.
FANCM	14	45644816	A	C	missense_variant	2	c.2859A>C	p.Lys953Asn	0.0031	0.0031	9	5		0.0011	0.0018	.	.	deleterious(0.688)	14.93	0.05	D	0.999	D	rs142864437	.	.	.	.
FANCM	14	45645144	A	G	missense_variant	3	c.3187A>G	p.Lys1063Glu	0.0003	0	1	0		.	.	.	.	neutral(0.060)	7.373	0.1	T	0.421	B	.	.	.	.	.
FANCM	14	45645150	T	G	missense_variant	3	c.3193T>G	p.Cys1065Gly	0.0003	0	1	0		.	.	0.0002	.	neutral(0.301)	2.486	0.03	D	0.267	B	rs201836469	.	.	.	.
FANCM	14	45645433	T	C	missense_variant	3	c.3476T>C	p.Leu1159Ser	0.0003	0	1	0		.	.	.	.	neutral(0.001)	0.007	0.74	T	0.013	B	.	.	.	.	.
FANCM	14	45645661	G	T	missense_variant	2	c.3704G>T	p.Gly1235Val	0	0.0005	0	1		8.30E-06	1.51E-05	.	.	deleterious(0.935)	18.79	0	D	1	D	rs146490925	.	.	.	.
FANCM	14	45645702	A	G	missense_variant	3	c.3745A>G	p.Thr1249Ala	0.0003	0	1	0		.	.	.	.	neutral(0.000)	0.004	0.91	T	0	B	.	COSM553235	1(lung)	.	.
FANCM	14	45645784	C	T	missense_variant	3	c.3827C>T	p.Ser1276Leu	0	0.0005	0	1		6.70E-05	4.54E-05	.	.	neutral(0.002)	0.061	0.41	T	0.004	B	.	COSM275090	1(large_intestine)	.	.
FANCM	14	45645826	C	T	missense_variant	3	c.3869C>T	p.Thr1290Ile	0.0003	0	1	0		2.48E-05	4.51E-05	.	.	neutral(0.056)	3.131	0.27	T	0.002	B	.	.	.	.	.
FANCM	14	45645855	G	T	stop_gained	1	c.3898G>T	p.Glu1300Ter	0	0.0005	0	1		.	.	.	.	0	12.64	0.78	T	.	.	.	.	.	.	.
FANCM	14	45645864	C	A	missense_variant	3	c.3907C>A	p.Gln1303Lys	0.0003	0	1	0		8.25E-06	1.50E-05	.	.	neutral(0.009)	2.981	0.32	T	0.483	P	rs369549739	.	.	.	.
FANCM	14	45645949	C	T	missense_variant	3	c.3992C>T	p.Pro1331Leu	0	0.0005	0	1		0.0001	0.0002	.	.	neutral(0.420)	11.32	0.03	D	0.289	B	rs149348098	.	.	.	.
FANCM	14	45645955	A	C	missense_variant	3	c.3998A>C	p.Gln1333Pro	0.0006	0.0005	2	1		9.20E-05	0.0002	.	.	neutral(0.038)	2.52	0.2	T	0.022	B	rs143681767	.	.	.	.
FANCM	14	45645978	C	G	missense_variant	3	c.4021C>G	p.Leu1341Val	0	0.0005	0	1		.	.	.	.	neutral(0.343)	10.39	0.3	T	0.996	D	.	.	.	.	.
FANCM	14	45646041	G	A	missense_variant	3	c.4084G>A	p.Asp1362Asn	0.0003	0	1	0		0.0002	0.0002	.	.	neutral(0.198)	12.61	0.45	T	0.915	P	rs199895244	.	.	.	.
FANCM	14	45650888	C	T	missense_variant	3	c.4366C>T	p.Arg1456Cys	0.0003	0.0005	1	1		0.0002	0.0003	.	.	neutral(0.261)	12.08	0.16	T	0.032	B	rs200360968	.	.	.	.
FANCM	14	45658342	A	C	missense_variant	3	c.5117A>C	p.Asn1706Thr	0	0.0005	0	1		2.48E-05	3.01E-05	.	.	neutral(0.003)	1.321	0.77	T	0.008	B	.	.	.	.	.
FANCM	14	45658402	C	T	missense_variant	3	c.5177C>T	p.Pro1726Leu	0.0003	0.0005	1	1		4.12E-05	6.00E-05	.	.	neutral(0.021)	9.94	0.18	T	0.039	B	.	.	.	.	.
FANCM	14	45665603	G	A	missense_variant	2	c.5569G>A	p.Val1857Met	0.0006	0.0017	2	3		0.0004	0.0005	0.0008	0.001	deleterious(0.935)	19.52	0	D	1	D	rs144008013	.	.	.	.

Gene	Chr	Position (GRCh37)	Ref	Alt	CONSEQUENCE	Class	c.DNA Change	Protein Change	Control_MAF	Case_MAF	ALT_Control	ALT_Cases	ALT_Ams	ExAC_Freq	ExAC_NFE	1000G_ALL	1000G_EUR	Condel	CADD_PHRED	SIFT_score	SIFT_pred	Polyphen2_HDIV_score	Polyphen2_HDIV_pred	dbSNP	COSMIC_ID	COSMIC_DIS	ClinVar_SIG	ClinVar_DIS
FANCM	14	45665613	G	A	missense_variant	2	c.5579G>A	p.Arg1860His	0.0003	0	1	0		2.47E-05	4.50E-05	.	.	deleterious(0.935)	26.4	0	D	1	D	.	COSM955827	1(large_intestine);1(endc	.	.
FANCM	14	45665637	A	G	missense_variant	2	c.5603A>G	p.Gln1868Arg	0	0.0005	0	1		1.65E-05	1.50E-05	.	.	deleterious(0.752)	16.69	0	D	0.999	D	.	.	.	.	.
FANCM	14	45667921	C	T	stop_gained	1	c.5791C>T	p.Arg1931Ter	0.0003	0.0021	1	4		0.0009	0.0009	0.0006	0.003	0	40	1	T	.	.	rs144567652	.	.	.	.
FANCM	14	45669207	T	C	missense_variant	3	c.6143T>C	p.Ile2048Thr	0.0003	0.0005	1	1		3.31E-05	6.02E-05	.	.	neutral(0.453)	10.92	0	D	0	B	rs150447576	.	.	.	.
RPS20	8	56986271	C	G	missense_variant	2	c.160G>C	p.Val54Leu	0	0.0005	0	1		8.24E-06	1.50E-05	.	.	deleterious(0.760)	19.8	0	D	0.99	D	.	.	.	.	.
PTPN12	7	77240335	A	G	missense_variant	3	c.911N>G	p.His304Arg	0	0.0005	0	1		0.0001	0.0002	.	.	neutral(0.023)	10.98	0.41	T	0	B	rs200238133	.	.	.	.
PTPN12	7	77256177	A	G	missense_variant	2	c.1181N>G	p.His394Arg	0.0003	0	1	0		.	.	.	.	deleterious(0.799)	21	0.78	T	0.999	D	.	.	.	.	.
PTPN12	7	77256234	C	T	missense_variant	3	c.1238N>T	p.Thr413Ile	0	0.0005	0	1		.	.	.	.	neutral(0.299)	9.173	0.01	D	0.04	B	.	.	.	.	.
PTPN12	7	77256365	A	C	missense_variant	2	c.1369N>C	p.Asn457His	0	0.001	0	2		1.66E-05	3.01E-05	.	.	deleterious(0.716)	17.21	0.04	D	0.97	D	rs145330429	.	.	.	.
PTPN12	7	77256375	A	G	missense_variant	3	c.1379N>G	p.His460Arg	0.0003	0	1	0		.	.	.	.	neutral(0.378)	17.85	0.08	T	0.99	D	.	.	.	.	.
PTPN12	7	77256453	A	G	missense_variant	3	c.1457N>G	p.Asn486Ser	0.0003	0	1	0		.	.	.	.	neutral(0.001)	0.003	0.89	T	0.002	B	.	.	.	.	.
PTPN12	7	77256502	A	T	missense_variant	3	c.1506N>T	p.Gln502His	0	0.0011	0	2		3.30E-05	6.00E-05	.	.	neutral(0.249)	3.617	0.3	T	0.001	B	.	.	.	.	.
PTPN12	7	77256528	C	T	missense_variant	3	c.1532N>T	p.Pro511Leu	0.0003	0	1	0		.	.	.	.	neutral(0.025)	0.64	0.67	T	0.007	B	.	.	.	.	.
PTPN12	7	77256562	G	T	missense_variant	2	c.1566N>T	p.Arg522Ser	0	0.0005	0	1		.	.	.	.	deleterious(0.771)	19.04	0.1	T	0.999	D	rs374516234	.	.	.	.
PTPN12	7	77256600	C	T	missense_variant	3	c.1604N>T	p.Thr535Met	0	0.0005	0	1		8.24E-06	1.50E-05	.	.	neutral(0.032)	5.044	0.23	T	0.004	B	.	.	.	.	.
PTPN12	7	77256672	A	G	missense_variant	3	c.1676N>G	p.Asn559Ser	0	0.0005	0	1		3.30E-05	1.50E-05	.	.	neutral(0.011)	0.015	0.85	T	0.001	B	rs61757751	.	.	.	.
PTPN12	7	77256872	A	G	missense_variant	2	c.1876N>G	p.Thr626Ala	0.0003	0	1	0		8.25E-06	1.50E-05	.	.	deleterious(0.853)	20.5	0.01	D	0.999	D	.	.	.	.	.
PTPN12	7	77261719	C	T	missense_variant	2	c.2051N>T	p.Ser684Leu	0.0006	0.0015	2	3		0.0009	0.0012	.	.	deleterious(0.889)	26.8	0	D	1	D	rs201001953	.	.	.	.
PTPN12	7	77267946	C	G	missense_variant	3	c.2179N>G	p.Pro727Ala	0.0003	0	1	0		.	.	.	.	neutral(0.001)	1.863	0.97	T	0.104	B	.	.	.	.	.
PTPN12	7	77256690	CTG	C	frameshift_variant	1	c.1695_1696delTG	p.Val566GlufsTer11	0.0003	0	1	0	0						35									
BUB1	2	111397347	C	G	missense_variant	3	c.3034G>C	p.Glu1012Gln	0	0.0005	0	1		.	.	.	.	neutral(0.447)	14.39	0.69	T	0.026	B	.	.	.	.	.
BUB1	2	111397415	A	G	missense_variant	2	c.2966T>C	p.Phe989Ser	0.0003	0	1	0		.	.	.	.	deleterious(0.906)	18.63	0.01	D	1	D	.	.	.	.	.
BUB1	2	111398700	A	G	missense_variant	2	c.2866T>C	p.Phe956Leu	0.0003	0	1	0		.	.	.	.	deleterious(0.911)	29.6	0.33	T	1	D	.	.	.	.	.
BUB1	2	111406857	C	T	stop_gained	1	c.2301G>A	p.Trp767Ter	0.0006	0	2	0		1.65E-05	3.00E-05	0.0002	0.001	0	39	0.24	T	.	.	rs201494812	.	.	untested	Malignant_melanoma
BUB1	2	111406882	C	T	missense_variant	3	c.2276G>A	p.Ser759Asn	0	0.0005	0	1		6.60E-05	0.0001	0.0002	.	neutral(0.005)	11.01	0.26	T	0.036	B	rs182377124	.	.	.	.
BUB1	2	111406904	A	G	missense_variant	2	c.2254T>C	p.Ser752Pro	0.0003	0	1	0		8.26E-06	1.50E-05	.	.	deleterious(0.892)	19.64	0.17	T	1	D	.	.	.	.	.
BUB1	2	111408213	C	T	missense_variant	3	c.2113G>A	p.Ala705Thr	0.0003	0	1	0		1.66E-05	1.50E-05	.	.	neutral(0.015)	9.661	0.7	T	0.002	B	.	.	.	.	.
BUB1	2	111414668	G	A	missense_variant	3	c.1643C>T	p.Thr548Ile	0.0006	0.0005	2	1		3.30E-05	4.50E-05	.	.	neutral(0.020)	7.683	0.24	T	0.001	B	rs376146753	.	.	.	.
BUB1	2	111416221	A	T	missense_variant	2	c.1375T>A	p.Ser459Thr	0	0.0005	0	1		1.65E-05	3.00E-05	.	.	deleterious(0.919)	26.2	0.37	T	1	D	.	.	.	.	.
BUB1	2	111416229	A	G	missense_variant	3	c.1367T>C	p.Val456Ala	0	0.0005	0	1		3.30E-05	4.50E-05	0.0002	.	neutral(0.311)	16.19	0.38	T	0.818	P	rs531352750	.	.	.	.
BUB1	2	111419297	G	T	missense_variant	2	c.1079C>A	p.Pro360His	0.0003	0	1	0		.	.	.	.	deleterious(0.616)	16.32	0.55	T	0.997	D	.	.	.	.	.
BUB1	2	111419307	C	T	missense_variant	3	c.1069G>A	p.Glu357Lys	0.0007	0	2	0		2.48E-05	3.01E-05	.	.	neutral(0.001)	3.291	0.79	T	0	B	.	.	.	.	.
BUB1	2	111425226	G	A	missense_variant	3	c.677C>T	p.Ala226Val	0.0044	0.0065	14	12		0.0019	0.0028	0.0012	0.006	neutral(0.409)	21.2	0.44	T	0.901	P	rs61730706	.	.	.	.
BUB1	2	111425402	A	G	missense_variant	3	c.592T>C	p.Ser198Pro	0.0003	0	1	0		.	.	.	.	neutral(0.059)	12.33	0.33	T	0.01	B	.	.	.	.	.
BUB1	2	111430329	C	T	missense_variant	3	c.331G>A	p.Ala111Thr	0.0003	0	1	0		.	.	.	.	neutral(0.003)	7.02	0.58	T	0.128	B	.	.	.	.	.
BUB1	2	111430340	C	T	missense_variant	3	c.320G>A	p.Gly107Glu	0	0.0006	0	1		4.95E-05	9.00E-05	.	.	neutral(0.001)	0.008	0.36	T	0.052	B	rs369472330	.	.	.	.
BUB1	2	111430353	T	C	missense_variant	3	c.307A>G	p.Ile103Val	0.0003	0	1	0		8.25E-05	3.00E-05	0.0002	.	neutral(0.005)	6.817	1	T	0	B	rs376910351	.	.	.	.
BUB3	10	124920001	A	G	missense_variant	2	c.496A>G	p.Asn166Asp	0.0003	0	1	0		.	.	.	.	deleterious(0.753)	33	0.1	T	0.99	D	.	.	.	.	.
BUB3	10	124921765	G	T	missense_variant	2	c.590G>T	p.Ser197Ile	0.0003	0	1	0		.	.	.	.	deleterious(0.850)	26.6	0	D	1	D	.	.	.	.	.
BUB3	10	124921827	A	T	stop_gained	1	c.652A>T	p.Lys218Ter	0.0006	0	2	0		.	.	.	.	0	37	1	T	.	.	.	.	.	.	.
BUB3	10	124922242	C	T	missense_variant	3	c.869C>T	p.Thr290Met	0	0.0005	0	1		1.65E-05	0	.	.	neutral(0.029)	8.721	0.2	T	0.008	B	rs151202304	.	.	.	.
BUB3	10	124922274	A	G	missense_variant	3	c.901A>G	p.Met301Val	0.0003	0	1	0		2.47E-05	4.50E-05	.	.	neutral(0.034)	5.416	0.16	T	0.005	B	rs140277834	.	.	.	.
RPS20	8	56985826	CAA	C	frameshift_variant	1	c.181_182delTT	p.Leu61GlufsTer11	0	0.0005	0	1						0	25									
BUB1	2	111415994	CTG	C	frameshift_variant	1	c.1507_1508delCA	p.Gln503ValfsTer15	0.0003	0	1	0						0	37									
BUB1	2	111427085	ATT	A	frameshift_variant	1	c.510_511delAA	p.Gln170HisfsTer11	0.0003	0	1	0		8.24E-06	0	.	.	0	22	.	.	.	.	.	.	.	.	.

Gene
Chr
Position (GRCh37)
Ref
Alt
CONSEQUENCE
Class
c.DNA Change
Protein Change
Control_MAF
Case_MAF
ALT_Control
ALT_Cases
ALT_Ams
ExAC_Freq
ExAC_NFE
1000G_ALL
1000G_EUR
Condel
CADD_PHRED
SIFT_score
SIFT_pred
Polyphen2_HDIV_score
Polyphen2_HDIV_pred
dbSNP
COSMIC_ID
COSMIC_DIS
ClinVar_SIG
ClinVar_DIS

Supplementary Figure 1: Forest plot of allelic odds ratio associated with *FANCM* p.Arg1931Ter (rs144567652) genotype and CRC. Studies [SMITH: original publication (2,207 cases, 2,176 controls)[1]; WES: whole-exome sequencing analyzed in this manuscript (863 cases, 1,604 controls); ENG: English Illumina Exome-BeadChip replication series (3,537 cases, 4,811 controls); SCOT: Scottish Illumina Exome-BeadChip replication series (2,015 cases, 1,981 controls)] were weighted according to the inverse of the variance of the log of the odds ratio (OR) calculated by unconditional logistic regression. Meta-analysis under a fixed-effects model was conducted using standard methods. Cochran's *Q* statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity were calculated. Horizontal lines indicate 95% confidence intervals (CIs). Boxes indicate OR point estimate; its area is proportional to the weight of the study. Diamond (and broken line) indicates overall summary estimate, with CI given by its width. Unbroken vertical line indicates null value (OR=1.0).