Identification of genetic variants for clinical management of familial colorectal tumors

Mev Dominguez-Valentin1*, Sigve Nakken1, Hélène Tubeuf2,3, Daniel Vodak1, Per Olaf Ekstrøm1, Anke M. Nissen4,5, Monika Morak4,5, Elke Holinski-Feder4,5, Alexandra Martins2, Pål Møller1,6,7 and Eivind Hovig1,8,9

Abstract

Background: The genetic mechanisms for families who meet the clinical criteria for Lynch syndrome (LS) but do not carry pathogenic variants in the mismatch repair (MMR) genes are still undetermined. We aimed to study the potential contribution of genes other than MMR genes to the biological and clinical characteristics of Norwegian families fulfilling Amsterdam (AMS) criteria or revised Bethesda guidelines.

Methods: The Hereditary Cancer Biobank of the Norwegian Radium Hospital was interrogated to identify individuals with a high risk of developing colorectal cancer (CRC) for whom no pathogenic variants in MMR genes had been found in routine diagnostic DNA sequencing. Forty-four cancer susceptibility genes were selected and analyzed by using our in-house designed TruSeq amplicon-based assay for targeted sequencing. RNA splicing- and protein-dedicated in silico analyses were performed for all variants of unknown significance (VUS). Variants predicted as likely to affect splicing were experimentally analyzed by resorting to minigene assays.

Results: We identified a patient who met the revised Bethesda guidelines and carried a likely pathogenic variant in CHEK2 (c.470 T > C, p.I157T). In addition, 25 unique VUS were identified in 18 individuals, of which 2 exonic variants (MAP3K1 c.764A > G and NOTCH3 c.5854G > A) were analyzed in the minigene splicing assay and found not to have an effect on RNA splicing.

Conclusions: Among high-risk CRC patients that fulfill the AMS criteria or revised Bethesda guidelines, targeted gene sequencing identified likely pathogenic variant and VUS in other genes than the MMR genes (CHEK2, NOTCH3 and MAP3K1). Our study suggests that the analysis of genes currently excluded from routine molecular diagnostic screens may confer cancer susceptibility.

Keywords: Lynch syndrome, Gene panel testing, CHEK2, RNA splicing mutations

Background

Heredity represents a major cause of colorectal cancer (CRC) with at least 20% of the cases estimated to develop due to genetic factors and about 5% being linked to inherited variants in cancer-predisposing genes [1–4]. Currently, patients with CRC are referred to germline mismatch repair (MMR) testing based on the identification of high-risk phenotypic features (i.e. early age of onset, family history, clinical criteria), but beyond microsatellite instability (MSI) and MMR immunohistochemistry (IHC) testing for Lynch syndrome (LS), no systematic approach to hereditary risk assessment exists [5].

LS is caused by a defective MMR system due to presence of germline defects in at least one of the MMR genes, MLH1, MSH2, MSH6, PMS2 or to deletions of the 3′ portion of the EPCAM gene [6]. LS is clinically classified according to the Amsterdam (AMS) criteria and/or the Bethesda guidelines, both relying in clinical information and family history. The Bethesda guidelines also take into account the MSI signature characteristic of MMR-deficient tumors [7–10]. LS patients have an increased lifetime risk of CRC (70–80%), endometrial cancer (50–60%), stomach cancer (13–19%), ovarian...
cancer (9–14%), cancers of the small intestine, the biliary tract and brain as well as carcinoma of the ureters and renal pelvis [11].

However, a high proportion of cases who meet the clinical criteria for LS (~60%) do not carry pathogenic variants in the MMR genes and have been reported as familial colorectal cancer type X (FCCTX) or Lynch-like syndrome (LLS) according to their MSI status [12–16]. The genetic mechanisms are undetermined in the majority of these families [14].

DNA sequencing (DNA-seq) studies using multigene panels have reported that as much as ~18% of patients diagnosed with CRC below the age of 50 years have pathogenic variants in several genes that are not traditionally associated with CRC (ATM, CHEK2, BRCA1, BRCA2, CDKN2A and PALB2) [5, 17]. Notably, there is a need to determine whether these variants contribute to hereditary CRC risk via the combination of low- and moderate-penetrance susceptibility alleles [5, 17, 18].

Given the high frequency and wide spectrum of pathogenic variants, it has been suggested that genetic counseling and testing with a multigene panel should be considered for all patients with early-onset CRC [17, 19–23]. Importantly, the identification of high-risk CRC patients is a major issue, because morbidity and mortality from CRC and extracolonic cancers in these patients and their relatives can be decreased by early screening and intensive surveillance [19, 24–26].

In an effort to discover inherited genetic variants that influence biological and clinical characteristics of familial CRC developed in unrelated high-risk patients, who previously tested negative for pathogenic variants in MMR genes, we examined 44 cancer associated genes using next generation sequencing (NGS), and applied minigene-based assay to analyze the impact of a subset of genetic variants on RNA splicing.

**Methods**

**Study population**

The Hereditary Cancer Biobank of the Norwegian Radium Hospital was used to identify unrelated high-risk CRC individuals from families that fulfilled the AMS criteria or the revised Bethesda guidelines [7–10, 27]. By the standard diagnostic clinical techniques, all study subjects were demonstrated not to carry pathogenic variants or large genomic rearrangements in MMR genes (MLH1, MSH2, MSH6 or PMS2).

Ethical approval for the study was granted by the Norwegian Data Inspectorate and Ethical Review Board (ref 2015/2382). All examined patients signed an informed consent for their participation in the study.

**Targeted sequencing**

Genomic DNA was isolated from peripheral blood samples and targeted sequencing was carried out using a TrueSeq amplicon based assay v.1.5 on a MiSeq apparatus, as previously described [28, 29]. The 44-gene panel used in this study includes genes associated with cancer predisposition as described in a prior study [28, 29].

**Sequencing data analysis**

Paired-end sequence reads were aligned to the human reference genome (build GRCh37) using the BWA-mem algorithm (v.0.7.8-r55) [30]. The initial sequence alignments were converted to BAM format and subsequently sorted and indexed with SAMtools (v.1.1) [30]. Genotyping of single nucleotide variants (SNV) and short indels was performed by GATK’s HaplotypeCaller. Filtering of raw genotype calls and assessment of callable regions/loci were done according to GATK’s best practice procedures, as described more detailed previously [28].

Variants were annotated using ANNOVAR (version November 2015) [31] and were queried against a range of variant databases and protein resources, namely dbSNP (build 147) [32], 1000 Genome Project phase3 [33], Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org, accessed August 2015) [34], Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org, accessed October 2017) [34], Norwegian Germline Variations Database (http://norgen-e.no/vcf-miner/, accessed October 2017), ClinVar (May 2016) [35], UniProt Knowledgebase (release March 2016) [36] and the Pfam protein domain database (v29, December 2015) [28, 37].

**Nomenclature and classification of genetic variants**

The nomenclature guidelines of the Human Genome Variation Society (HGVS) were used to describe the detected genetic variants [38]. The recurrence of the identified variants was established by interrogating four databases (in their latest releases as of November 2016): the Leiden Open Variation Database (LOVD), the Universal Mutation Database (UMD), ClinVar and the Human Gene Mutation Database (HGMD). The variants were classified according to the 5-tier classification system into the following categories: class 5 (pathogenic), class 4 (likely pathogenic), class 3 (uncertain variants or variants of unknown significance, VUS), class 2 (likely not pathogenic) and class 1 (not pathogenic) [3].

**In silico analyses of VUS**

Two types of bioinformatics methods were used to predict the impact of selected variants on RNA splicing. First, we used MaxEntScan (MES) and SSF-like (SSFL) to predict variant-induced alterations in 3′ and 5′ splice site strength, as described by Houdayer et al. 2012 [39],...
except that here both algorithms were interrogated by using the integrated software tool Alamut Batch version 1.5, (Interactive Biosoftware, http://www.interactive-biosoftware.com). For prediction of variant-induced impact on exonic splicing regulatory elements (ESR), we resorted to ΔtESRseq- [40], ΔHZei- [41], and SPANR-based [42] as described by Soukarieh et al. [43]. Score differences (Δ) between variant and wild-type (WT) cases were taken as proxies for assessing the probability of a splicing defect. More precisely, we considered that a variant mapping at a splice site was susceptible of negatively impacting exon inclusion if ΔMES≥15% and ΔSSFL≥5% [39], whereas an exonic variant located outside the splice site was considered as a probable inducer of exon skipping if negative Δ scores (below the thresholds described below) were provided by all the 3 ESR-dedicated in silico tools. We chose the following thresholds: −0.5 for ΔtESRseq-, −10 for ΔHZei-, and < −0.5 for SPANR-based scores. In addition, we evaluated the possibility of variant-induced de novo splice sites by taking into consideration local changes in MES and SSFL scores. In this case, we considered that variants located outside the splice sites were susceptible of creating a competing splice site if local MES scores were equal to or greater than those of the corresponding reference splice site for the same exon.

In silico protein impact predictions of missense variants were performed with Align-GVGD (the VUS were predicted as deleterious when the values were from C35 or higher), SIFT, and MAPP using Alamut Batch version 1.4.4 (Interactive Biosoftware) and additionally with PolyPhen-2 and MutationTaster [44–48].

Cell-based minigene splicing assays
In order to determine the impact of selected exonic variants on splicing, we performed functional assays based on the comparative analysis of the splicing pattern of WT and mutant reporter minigenes, as follows. First, genomic regions containing the exon of interest (internal exons only) and at least 150 nucleotides of the flanking introns were amplified by PCR [49] using patients’ DNA as template and primers indicated in Additional file 1: Table S1. Next, representative minigenes were created by inserting the PCR-amplified fragments into a previously linearized pCAS2 vector [43]. All constructs were sequenced to ensure that no unwanted mutations had been introduced into the inserted fragments during PCR or cloning. Then, WT and mutant minigenes were transfected into HeLa cells grown in 12-well plates (at ~70% confluence) using the FuGENE 6 transfection reagent (Roche Applied Science). Twenty-four hours later, total RNA was extracted using the NucleoSpin RNA II kit (Macherey Nagel) and, the minigenes’ transcripts were analyzed by semi-quantitative RT-PCR using the OneStep RT-PCR kit (Qiagen), as previously described [43]. The sequences of the RT-PCR primers are shown in Additional file 1: Table S1. Later, RT-PCR products were separated by electrophoresis on 2.5% agarose gel containing EtBr and visualized by exposure to UV light under saturating conditions using the Gel Doc XR image acquisition system (Bio-Rad), followed by gel-purification and Sanger sequencing for proper identification of the minigenes’ transcripts. Finally, splicing events were quantitated by performing equivalent fluorescent RT-PCR reactions followed by capillary electrophoresis on an automated sequencer (Applied Biosystems), and computational analysis by using the GeneMapper v5.0 software (Applied Biosystems).

Results
Clinical characteristics and family history
Upon querying the Hereditary Cancer Biobank of the Norwegian Radium Hospital for cases that fulfill the AMS and/or the revised Bethesda guidelines, we identified 34 unrelated potential high-risk CRC individuals who did not carry pathogenic variants in MMR genes. The median age at first CRC diagnosis was 51.5 years (range: 34–86 years).

Pedigree information showed that 13 (38%) families fulfilled the AMS I and/or II criteria and the revised Bethesda guidelines while 21 (62%) met the revised Bethesda guidelines only (Table 1). Fifteen (44%) patients had tumors with MSI and/or MMR IHC data available, of which 2 (13%) were MSI-high and/or MMR deficient. Clinical, family and tumor data information is detailed in Table 1.

Germline findings
Given that the families that fulfilled the AMS criteria and/or the Bethesda guidelines did not carry pathogenic variants in the MMR genes, we hypothesized that other genes could be implicated in the genetic determinism of these phenotypes.

In order to pursue this hypothesis, we collected DNA samples from all probands and performed high-throughput sequencing of a panel of 44 cancer-associated genes. For the 34 samples, mean depth of coverage ranged from 127 to 507 with the fraction of target bases with coverage ≥25 ranging from 80% to 93. The NGS results revealed that each individual carried an average of 26 SNV (between 19 and 33 per individual) in the set of 44 cancer susceptibility genes, most of which were common polymorphisms (allele frequency ≥1% in the general population) according to the ExAC database, and some being classified as benign or likely benign (class 1 or class 2) according to either ClinVar or the American College of Medical Genetics and Genomics (ACMG) guidelines [35, 50] (Table 2).
Importantly, we identified a likely pathogenic variant in a moderate-penetrance gene (CHEK2 c.470 T > C, p.I157T) in a female patient diagnosed with colon cancer at 42 years, melanoma at 44 years and BC at 57 years with a proficient IHC MMR profile and fulfilling the revised Bethesda guidelines (Patient 19,609) (Table 1).

The CHEK2 c.470 T > C has been classified as pathogenic according to the ACMG guidelines [51], and has a lower allele frequency (1.89*10⁻³) in the Norwegian population, compared to the non-Finnish European population (5.4*10⁻³) (http://norgene.no/vcf-miner/ and gnomAD database, respectively) [34, 35, 50]. The variant

### Table 1 Summary of International Classification of Diseases, 9th Revision (ICD9), gender, age at diagnosis, clinical criteria and tumor molecular characteristics of the familial CRC families

| Patient_ID | Gender | ICD9 diagnosis (age) | AMS criteria | Revised Bethesda | Tumor molecular characteristics |
|------------|--------|----------------------|--------------|------------------|-------------------------------|
| 3222       | F      | CC (54), Hyperplastic polyp (55/61/62/63/65), BC (70) | 0            | Y                | MMR IHC proficient            |
| 3308       | F      | CC (43), BC (51/52) | 0            | Y                | MMR IHC proficient            |
| 3387       | F      | BC (40), OC (70), CC (80) | 0            | Y                | MMR IHC proficient            |
| 3426       | M      | MM (39) | I & II | Y                | na                           |
| 4932       | F      | CC (34), EC (40), Hyperplastic polyp (43), BT (46) | I & II | Y                | na                           |
| 5324       | F      | M (52), CC (59), SMC (na), BC (72) | 0            | Y                | na                           |
| 6174       | F      | Hyperplastic polyp (63/67), BC (65) | I & II | Y                | MMR IHC proficient            |
| 6977       | F      | TC (66) | 0            | Y                | MMR IHC proficient            |
| 9876       | F      | M (45), BC (54) | 0            | Y                | na                           |
| 9998       | F      | Hyperplastic polyp (45), CC (45) | II | Y                | MMR IHC proficient            |
| 10,675     | F      | BC (51), Hyperplastic polyp (59), TC (60) | II | Y                | na                           |
| 12,954     | F      | Hyperplastic polyp (69), ML (70) | II | Y                | na                           |
| 13,072     | M      | Hyperplastic polyp (63/64/65), CC (65/67) | 0            | Y                | na                           |
| 14,930     | F      | Hyperplastic polyp (86), CC (86) | 0            | Y                | MMR IHC proficient            |
| 18,843     | F      | BC (44), CC (49), SMC (na) | 0            | Y                | na                           |
| 19,411     | M      | PC (70) | 0            | Y                | MSH6 IHC deficient            |
| 19,673     | F      | BC (40/42) | II | Y                | na                           |
| 20,612     | F      | Hyperplastic polyp (59/65), EC (70) | 0            | Y                | na                           |
| 21,368     | F      | OC (62) | 0            | Y                | na                           |
| 22,295     | F      | Hyperplastic polyp (53), M (58) | 0            | Y                | na                           |
| 23,761     | F      | Hyperplastic polyp (40/42/44), BC (50) | 0            | Y                | na                           |
| 23,910     | F      | M (43), Hyperplastic polyp (49), BC (63), BT (63) | 0            | Y                | na                           |
| 24,140     | F      | CC (45/67), BC (56) | 0            | Y                | na                           |
| 24,447     | F      | BC (57/66), CC (66) | 0            | Y                | MLH1/PMS2 IHC deficient and MSI |
| 11,705     | F      | THC (53), KC (53/63) | II | Y                | MMR IHC proficient            |
| 12,673     | F      | OC (23), SMC (36), RC (62) | II | Y                | na                           |
| 13,393     | M      | RC (48), CST (58) | I & II | Y                | MMR IHC proficient and MSS    |
| 14,963     | F      | Hyperplastic polyp (69), BC (62) | 0            | Y                | na                           |
| 19609a     | F      | CC (42), M (44), BC (57) | 0            | Y                | MMR IHC proficient            |
| 22,953     | F      | BC (53) | II | Y                | na                           |
| 24,789     | F      | CC (43), RC (65), BC (72) | I & II | Y                | MMR IHC proficient and MSS    |
| 25,167     | M      | CC (55) | I & II | Y                | MMR IHC proficient            |
| 5597       | M      | Hyperplastic polyp (53/54), SC (55), BC (62) | 0            | Y                | MMR IHC proficient            |
| 8913       | F      | Hyperplastic polyp (59), BC (61), TC (69) | 0            | Y                | na                           |

**CRC** colorectal, ICD9 diagnosis International Classification of Diseases, 9th Revision, CC colon cancer, BC breast cancer, AMS Amsterdam criteria, 0 not fulfill the AMS criteria, Y yes, MMR mismatch repair, IHC immunohistochemistry, MSI microsatellite instability, MSS microsatellite stable, na not available, OC ovary cancer, MM multiple myeloma, EC endometrial cancer, BT brain tumor, M melanoma, SMC other malignant of the skin, TC trachea, bronchus, lung cancer, ML malignant neoplasms of lymphoid, PC prostate cancer, THC thyroid cancer, KC kidney cancer, RC rectum cancer, CST malignant neoplasm of connective and soft tissue, SC stomach cancer

*Patient carrying CHEK2 c.470 T > C, p.I157T*
| Patient ID | VUS | Benign or Likely Benign variants (Class 1 or 2) | Polymorphisms | Total variants/patient |
|------------|-----|-----------------------------------------------|---------------|-----------------------|
| 3222       | NOTCH3 | ATN NM_000051: c.5071A > C, p.S1691R (rs1800059) | rs459552, rs659243, rs2240308, rs1799966, rs16942, rs16941, rs169547, rs4986764, rs1805107, rs506504, rs832582, rs5868032, rs1042821, rs3210484, rs1044009, rs152451, rs2228006, rs1805321, rs4796033, rs1042522, rs861539, rs13125836 | 28 |
| 3308       | NBN | BARD1 NM_000465: c.1075_1095del, p.L359-P365delLPECSSP (rs28997575) | rs459552, rs659243, rs11528010, rs1444848, rs169547, rs4986764, rs1805107, rs506504, rs702689, rs832582, rs5868032, rs1799977, rs1042821, rs3210484, rs1805794, rs1044009, rs2228006, rs5744934, rs5744751, rs4796033, rs1042522 | 24 |
| 3387       | NA | CDKN2A N M_000077: c.442G > A, p.A148T (rs3731249) | rs459552, rs659243, rs1801516, rs2240308, rs2070094, rs2229571, rs11528010, rs1444848, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs702689, rs832582, rs5868032, rs3210489, rs1044009, rs2228006, rs1805321, rs5744934, rs1042522 | 23 |
| 3426       | NA | NA | rs459552, rs659243, rs1801516, rs2240308, rs2070094, rs2229571, rs1048108, rs11528010, rs1444848, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs702689, rs832582, rs5868032, rs1805794, rs1044009, rs2228006, rs1805321, rs5744934, rs1042522 | 23 |
| 4932       | NOTCH3 | NA | rs459552, rs659243, rs1801516, rs2240308, rs2229571, rs11528010, rs1799966, rs16942, rs16941, rs169547, rs1805107, rs506504, rs12642536, rs702689, rs832582, rs5868032, rs3210489, rs1805794, rs1044009, rs2228006, rs1805321, rs861539 | 26 |
| 5324       | NA | NA | rs459552, rs659243, rs1801516, rs2240308, rs2229571, rs11528010, rs1799966, rs16942, rs16941, rs169547, rs1805107, rs506504, rs12642536, rs702689, rs832582, rs5868032, rs3210489, rs1805794, rs1044009, rs2228006, rs1805321, rs861539 | 19 |
| Patient_ID | VUS (Class 3) | Benign or Likely Benign variants (Class 1 or 2) | Polymorphisms | Total variants/patient |
|------------|--------------|-----------------------------------------------|---------------|-----------------------|
| 6174       | na           | PMS2 NM_000535: c.1454C > A, p.T485K (rs1805323) | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs144848, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs12642536, rs832582, rs1799977, rs3219489, rs1044009, rs2228006, rs1805321, rs5744934 |
| 6977       | na           | BARDO1 NM_000465: c.1075_1095del, p.L359-P365delPECSSP (rs28997575) | rs459552, rs659243, rs1805156, rs2240308, rs1799966, rs16942, rs16941, rs799917, rs4986850, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs12642536, rs702689, rs832582, rs5868032, rs1799977, rs3219484, rs1805794, rs1044009, rs2228006, rs1805321, rs1042522, rs861539 |
| 9876       | PSMC3IP NM_016556: c.136G > A, p.V46M (rs757057684) RADS1B NM_133509: c.1063G > A, p.A355T (rs61758785) | RAD51D NM_002878: c.698A > G, p.E233G (rs28363284) | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs169547, rs4986764, rs1805107, rs506504, rs702689, rs832582, rs5868032, rs1799977, rs1044009, rs2228006, rs5744934, rs1042522, rs861539, rs28908468 |
| 9998       | na           | MSH6 NM_000179: c.2633 T > C, p.V878A (rs2020912) | rs459552, rs659243, rs1805156, rs2240308, rs2070094, rs2229571, rs1048108, rs169547, rs4986764, rs1805107, rs506504, rs702689, rs832582, rs5868032, rs1799977, rs1044009, rs2228006, rs1805321, rs5744751, rs1042522 |
| 10675      | na           | PMS2 NM_000535: c.1531A > G, p.T511A (rs2228007) | rs459552, rs659243, rs1805156, rs2070094, rs2229571, rs11528010, rs1799966, rs16942, rs16941, rs799917, rs1044009, rs2228006, rs1805321, rs5744751, rs1042522 |
| Patient ID | VUS (Class 3) | Benign or Likely Benign variants (Class 1 or 2) | Polymorphisms | Total variants/patient |
|------------|---------------|-----------------------------------------------|---------------|-----------------------|
| 13,072     | PMS2 NM_000535: c.1789A > T, p.T597S (rs1805318) | rs459552, rs659243, rs2240308, rs2229571, rs11528010, rs1799966, rs16942, rs16941, rs799917, rs144848, rs169547, rs4986764, rs1805107, rs506504, rs21642536, rs702689, rs832582, rs5868032, rs1799977, rs1805794, rs1044009, rs2228006, rs1805321, rs1726801, rs799917, rs4986850, rs1042522 | 22 |
| 14,930     | PALB2 NM_024675: c.925A > G, p.I309V (rs3809683) | rs459552, rs659243, rs1801516, rs2240308, rs2070094, rs2229571, rs11528010, rs1799966, rs16942, rs16941, rs799917, rs144848, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs12642536, rs702689, rs832582, rs5868032, rs1799977, rs1805794, rs1044009, rs2228006, rs5744751, rs1042522 | 30 |
| 18,843     | BRCA1 NM_007300: c.3119G > A, p.S1040N (rs4986852) | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs11528010, rs1799966, rs16942, rs16941, rs799917, rs4986850, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs12642536, rs702689, rs832582, rs5868032, rs1799977, rs3210498, rs1805794, rs1044009, rs152451, rs2228006, rs1805321, rs5744934, rs1042522 | 26 |
| 19,411     | BARD1 NM_000465: c.1972C > T, p.R658C (rs3738888) PALB2 NM_024675: c.1010T > A | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs1799966, rs16942, rs16941, rs799917, rs4986850, rs861539 | 28 |
| Patient ID | VUS (Class 3) | Benign or Likely Benign variants (Class 1 or 2) | Polymorphisms | Total variants/patient |
|-----------|---------------|-----------------------------------------------|----------------|------------------------|
| 19,673    | AXIN2 NM_004655: c.344A > G, p.N115S (rs370257532) | C, p.L337S (rs45494092) | rs144848, rs169547, rs1805107, rs506504, rs1126497, rs12642536, rs702689, rs832582, rs5868032, rs1805794, 1,044,009, rs2228006, rs1805321, rs5744751, rs1042522, rs861539 | 19,673 |
| 20,612    | na            | CDKN2A NM_000077: c.442G > A, p.A148T (rs3731249) | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs11528010, rs1799966, rs16942, rs16941, rs799917, rs144848, rs169547, rs4986764, rs1805107, rs506504, rs702689, rs832582, rs5868032, rs1042821, rs1805794, rs1044009, rs2228006, rs1805321, rs4796033, rs1042522, rs861539, rs3218536 | 20,612 |
| 21,368    | MAP3K1 NM_005921: c.764A > G, p.N255S (rs56069227) | ATM NM_000051: c.2572 T > C, p.F858L (rs1800056) | rs459552, rs659243, rs2240308, rs2229571, rs11528010, rs1799950, rs169547, rs1805107, rs506504, rs12642536, rs702689, rs5868032, rs1799977, rs1805794, rs1044009, rs2228006, rs1805321, rs10254120, rs5744934, rs4796033, rs1042522, rs861539 | 21,368 |
| 22,295    | na            | BRCA1 NM_007300: c.5019G > A, p.M1673I (rs1799967) | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs1048108, rs1799966, rs16942, rs16941, rs799917, rs144848, rs169547, rs1805107, rs506504, rs12642536, rs702689, rs832582, rs5868032, rs1799977, rs1044009, rs2228006, rs1805321, rs5744934, rs4796033, rs1042522 | 22,295 |
| 23,761    | na            | na | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs11528010, rs1799966, rs16942, | 23,761 |
| Patient ID | VUS (Class 3) | Benign or Likely Benign variants (Class 1 or 2) | Polymorphisms | Total variants/patient |
|------------|--------------|-----------------------------------------------|--------------|-----------------------|
| 23,910     | na           | BRCA2 NM_000059: c.6100C > T, p.R2034C (rs1799954) MSH2 NM_000251: c.965G > A, p.G322D (rs4987188) MSH6 NM_000179: c.2633 T > C, p.V878A (rs2020912) PALB2 NM_024675: c.2794G > A, p.V932 M (rs45624036) PMS2 NM_000535: c.1454C > A, p.T485 K (rs1805323) BARD1 NM_000465: c.1670G > C, p.C557S (rs28997576) | rs16941, rs799917, rs1799950, rs144848, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs702689, rs832582, rs5868032, rs1799977, rs1044009, rs2228006, rs1805321, rs1802683, rs1042522 | 28 |
| 24,140     | na           | BRCA1 NM_007300: c.5019G > A, p.M1673I (rs1799967) PMS2 NM_000535: c.1531A > G, p.T511A (rs2228007) | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs11528010, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs702689, rs832582, rs5868032, rs1799977, rs10442821, rs1805794, rs2228006, rs10254120, rs1042522, rs13125836, rs3218536 | 29 |
| 24,447     | CHEK2 NM_007194: c.74 T > C, p.V25A (rs587780188) | NOTCH3 NM_000435: c.3399C > A, p.H1133Q (rs112197217) | rs459552, rs659243, rs2240308, rs2070094, rs2229571, rs1048108, rs1799966, rs16942, rs16941, rs799917, rs444848, rs169547, rs1805107, rs506504, rs1126497, rs702689, rs832582, rs5868032, rs1799977, rs1042821, rs3219489, rs1805794, rs1044009, rs2228006, rs10254120, rs5744751, rs1042522 | 24 |
| 11,705     | ATM NM_000051: c.4375G > A, p.G1459R (rs145667735) MSH2 NM_000251: c.1284C > G, p.H428Q (rs776034412) | MSH2 NM_000251: c.965G > A, p.G322D (rs4987188) PMS2 NM_000535: c.1454C > A, p.T485 K (rs1805323) | rs459552, rs659243, rs1801516, rs2240308, rs2070094, rs2229571, rs1799966, rs16942, rs16941, rs799917, rs444848, rs169547, rs1805107, rs506504, rs1126497, rs702689, rs832582, rs5868032, rs1805794, rs1044009, rs2228006, rs1805321, rs1042522 | 27 |
| 12,673     | na           | ATM NM_000051: c.2572 T > C, | rs459552, rs659243, rs2240308, rs2070094, | 26 |
| Patient ID | VUS (Class 3) | Benign or Likely Benign variants (Class 1 or 2) | Polymorphisms | Total variants/patient |
|------------|---------------|-----------------------------------------------|---------------|-----------------------|
| 13,393     | NBN NM_002485.4: c.643C > T, p.R215W (rs34767364) | p.F858 L (rs1800056) PMS2 NM_000535: c.1454C > A, p.T485 K (rs1805323) | rs2229571, rs1048108, rs1799950, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs1246236, rs702689, rs32582, rs586032, rs1799977, rs1044009, rs2228006, rs1805321, rs1802683, rs4796033, rs1042522, rs861539 | 13,393 |
| 14,963     | PALB2 NM_024675: c.232G > A, p.V78I (rs515726085) | BARD1 NM_000465: c.1972C > T, p.R658C (rs3738888) BRIP1 NM_032043: c.577G > A, p.V193I (rs4988346) PMS2 NM_000535: c.1531A > G, p.T511A (rs2228007) ATM NM_000051: c.4258C > T, p.L1420F (rs1800058) NOTCH3 NM_000435: c.3058G > C, p.A1020P (rs35769976) NOTCH3 NM_000435: c.3547G > A, p.V1183 M (rs10408766) | rs459552, rs659243, rs2229571, rs1799966, rs16942, rs16941, rs799917, rs169547, rs4986764, rs1805107, rs506504, rs702689, rs832582, rs586032, rs1805794, rs1044009, rs2228006, rs1805321, rs1042522 | 14,963 |
| 19,609     | na            | BRCA2 NM_000059: c.4258G > T, p.D1420Y (rs28897727) PCLE NM_006231: c.2083 T > A, p.F695I (rs5744799) | rs459552, rs659243, rs2229571, rs1799966, rs16942, rs16941, rs799917, rs144848, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs12462536, rs702689, rs586032, rs1799977, rs3219489, rs1805794, rs1044009, rs2228006, rs1805321, rs5744751, rs1042522, rs861539, rs3218356 | 19,609 |
| 22,953     | NOTCH3 NM_000435: c.5208G > C, p.E1736D (rs200331646) MSH2 NM_0000251: c.128A > G, p.Y43C (rs17217723) RAD51B NM_133510: c.515 T > G, p.L172 W (rs34094401) | BRCA2 NM_000059: c.2971A > G, p.N991D (rs1799944) CDH1 NM_004360: c.1774G > A, p.A592T (rs35187787) | rs459552, rs659243, rs2229571, rs1799966, rs16942, rs16941, rs799917, rs3219489, rs1805794, rs1044009, rs2228006, rs1805321, rs5744934, rs1042522, rs13125836 | 22,953 |
| Patient ID | VUS (Class 3) | Benign or Likely Benign variants (Class 1 or 2) | Polymorphisms | Total variants/patient |
|------------|--------------|-----------------------------------------------|----------------|-----------------------|
| 24,789     | APC NM_001127510: c.4334C > T, p.T1445I (rs760683468) PALB2 NM_024675: c.1250 C > A, p.S417Y (rs45510998) | BARD1 NM_000465: c.1972C > T, p.R658C (rs3738888) PALB2 NM_024675: c.2993G > A, p.G998E (rs45551636) PALB2 NM_024675: c.2014G > C, p.E672Q (rs45532440) POLE NM_006231: c.776G > A, p.R259H (rs61732929) NOTCH3 NM_000435: c.3396C > A, p.H1133Q (rs112197217) | rs1044009, rs2228006, rs1805321, rs1726801 | 30 |
| 25,167     | NBN NM_002485.4: c.643C > T, p.R215W (rs34767364) | ATM NM_000051: c.2119T > C, p.S707P (rs4986761) | rs459552, rs659243, rs2240308, rs2229571, rs1799966, rs16942, rs16941, rs799917, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs12642536, rs702689, rs832582, rs5868032, rs1799977, rs3210489, rs1044009, rs152451, rs2228006, rs1042522 | 22 |
| 5597       | MAP3K1 NM_005921: c.2816C > G, p.S939C (rs45556841) | ATM NM_000051: c.5071A > C, p.S1691R (rs1800059) | rs459552, rs659243, rs1805161, rs2070094, rs2229571, rs11528010, rs169547, rs4986764, rs1805107, rs506504, rs12642536, rs702689, rs832582, rs5868032, rs1799977, rs1805794, rs2228006, rs1805321, rs1726801, rs4796033, rs1042522 | 29 |
| 8913       | RAD51B NM_133509: c.1603G > A, p.A355T (rs61758785) EPCAM NM_002354: c.267G > C, p.Q89H (rs146480420) | PM152 NM_000532: c.1454C > A, p.T485K (rs1805323) | rs459552, rs659243, rs2240308, rs2229571, rs1408108, rs11528010, rs1799966, rs16942, rs16941, rs799917, rs144848, rs169547, rs4986764, rs1805107, rs506504, rs1126497, rs12642536, rs702689, rs832582, rs5868032, rs1044009, rs2228006, rs5744751, rs1042522, rs861359, rs13125836 | 30 |

* Recently classified as Benign by ACMG Guidelines, 2015

* Patient ID carrying CHEK2 c.470 T > C, p.I157T
is reported in ClinVar as “conflicting interpretations of pathogenicity, risk factor” (Variation ID: 5591). When the revised Bethesda guidelines were considered, the mutation detection rate was thus 4.8% (1/21).

Overall, 25 unique VUS were found in 18 out of the 34 patients (Table 2). The detected VUS were distributed among 17 different genes: MAP3K1 (in 2 patients), NBN (in 3 patients), NOTCH3 (in 3 patients), RAD51B (in 3 patients), MSH2 (in 2 patients), PALB2 (in 2 patients), POLE (in 2 patients) and the remaining were found in APC, ATM, AXIN2, BRCA1, CHEK2, EPCAM, MSH6, MUTYH, RAD51C and STK11 (Table 2). The minor allele frequency (MAF) values of these variants were very low or no frequency data have been reported.

**Protein and splicing-dedicated in silico analyses**

The 25 unique VUS were analyzed by using five in silico prediction tools with different underlying algorithms to estimate the impact of the variants on the structure and function of the corresponding proteins.

Concordances between the 5 prediction tools were found for 2 out of the 25 VUS, suggesting a potentially damaging effect on protein level for the variants: MUTYH c.812G>A (p.R271Q) and MUTYH damaging effect on protein level for the variants: found for 2 out of the 25 VUS, suggesting a potentially function of the corresponding proteins.

The 25 unique VUS were analyzed by using five in silico analyses in our study do not fully reproduce the splicing pattern of the minigenes’ transcripts. These data thus disagree with the in silico predictions and suggest that either the exon 32 of NOTCH3 and the exon 3 of MAP3K1 are refractory to splicing mutations (the predictions thus being incorrect) or that the minigenes used in our study do not fully reproduce the splicing pattern of the mutant exons in NOTCH3 and MAP3K1 bona fide transcripts (the predictions being eventually correct). Complementary studies using RNA from NOTCH3 c.5854G>A and MAP3K1 c.764A>G carriers need to be performed to verify the pertinence of these results.

**Discussion**

The major unexpected finding in our Norwegian high-risk CRC cohort was the detection of a likely pathogenic variant in CHEK2 (c.470 T>C, p.I157T), a moderate-penetrance gene not traditionally associated with CRC, in an individual with a LS-evocative personal/family history and a high number of Class 3 variants in BC- and CRC-associated genes. Interestingly, the CHEK2 (c.470 T>C, p.I157T) has an allele frequency of 1.89*10^-3 in the Norwegian population (http://norgen.e.no/vcf-miner/), and is reported in ClinVar as having conflicting interpretations of pathogenicity/being a risk factor (Variation ID: 5591). Importantly, there is no systematic classification for most of the genetic variants found by NGS, and, in more general terms, the impact of low- to moderate-penetrance pathogenic variants with respect to clinical management is not fully understood [52]. Co-segregation or case-control studies for further evaluation will be key in understanding whether such germline variant may have a modifying effect, since we do not yet have evidence-based guidelines for the majority of these genes.

On the other hand, CHEK2 germline variants have been described to confer an elevated risk of BC (relative risk = 3.0) [53]. However, the presence of pathogenic variants in CHEK2 is not frequently associated with cancer in high-risk BC families, prompting speculation that there may be several low-penetrance or moderate-penetrance BC risk genes segregating independently
Table 3 In silico data obtained for the variants of unknown significance (VUS) identified in our study of familial CRC individuals

| Patient_ID | Genomic position (GRCh37) | Gene | Exon | Nucleotide change (cNomen) | Protein change (pNomen) | Reference splice site-dedicated analyses | Cryptic splice site-dedicated analyses | ESR-dedicated analyses | Protein-dedicated analyses |
|------------|---------------------------|------|------|---------------------------|-------------------------|-----------------------------------------|----------------------------------------|-----------------------|-------------------------|
| 3222       | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |
| 3222       | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |
| 3308       | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |
| 4932       | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |
| 5597       | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |
| 9876       | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |
| 9876       | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |
| 129        | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |
| 130        | chr19:195273C>T           | NOT  | CH3  | c.585G>A                   | p.V1015M                | 39_39WT                                  | 11.5 11.5 0                           | 89.2 89.2 0           | - - -                   |

**Dominguez-Valentin et al. BMC Medical Genetics (2018) 19:26**
| Patient ID | Genomic position (GRCh37) | Gene Exon | Nucleotide change (cNomen) | Predicted protein change (pNomen) | Reference splice site-dedicated analyses | Cryptic splice site-dedicated analyses | ISR-dedicated analyses | Protein-dedicated analyses |
|------------|---------------------------|-----------|-----------------------------|----------------------------------|-------------------------------------------|-------------------------------------|-------------------------|---------------------------|
| 188        | chr 248 027317 G>A         | MSH6 4**  | c.219 G>A                   | p.R73 Q                        | -978 5'                                   | 89 89 0                             | 81.6 81.6 0              | -0.80 5417 -13.15 0 C0 Tolerated good benign disease causing |
| 19         | chr 176 3554895 T>C        | A0N2 2   | c.344 A>G                    | p.N113                        | 460 3'                                   | 11.1 11.1 0                          | 93.3 93.3 0              | -0.16 2849 -2.19 0-0.05 C0 Tolerated good benign disease causing |
| 21         | chr 368 55672 A>G          | MAP 3     | c.764 A>G                    | p.N255                       | -71 5'                                   | 75 75 0                              | 78.5 78.5 0              | 4.7 8.8 -1.18 6.7 0.04 C0 Tolerated good benign polymorphism |
| 24         | chr 447 9130 636 A>G       | CHEK2 2   | c.74 T>C                     | p.V25A                        | 80 3'                                   | 1.7 1.7 0                            | 85.5 85.5 0              | -0.1 66 36.7 0.07 C0 Tolerated good benign polymorphism |
| 11         | chr 705 1110 8160 467 G>A  | ATM 29    | c.437 G>A                    | p.G14 5R                     | -62 5'                                   | 89 89 0                              | 87.5 87.5 0              | 0.5 52 0.26 -27.26 0.09 C15 Deleterious bad probably damaging disease causing |
| 133        | chr 93 8_90 983 460 G>A    | MSH2 8   | c.128 4C>G                   | p.H4 28Q                     | 8 3'                                   | 10.1 10.1 0                         | 87.3 87.3 0              | 0.7 0.3 22.05 0.05 C0 Tolerated good possibly damaging disease causing |
| 149        | chr 63 162 364 7665 C>T    | A0N2 4**  | c.232 G>A                    | p.V78I                        | 21 3'                                   | 100 100 0                           | 90.3 90.3 0              | 0.7 0.8 0.09 0.09 0.8 C0 Tolerated good polymorphism |
| 229        | chr 53 191 527 8214 C>G    | NAF 29    | c.5208 G>C                   | p.E17 360                     | 9 3'                                   | 78 78 0                             | 84.2 84.2 0              | 0.00 50 0.09 0.22 C35 Deleterious bad benign disease causing |
Table 3 In silico data obtained for the variants of unknown significance (VUS) identified in our study of familial CRC individuals (Continued)

| Patient_ID | Genomic position (GRCh37) | Gene | Exon | Nucleotide change (chrom) | Predicted protein change (pNumber) | Reference splice site-dedicated analyses | Cryptic splice site-dedicated analyses | Protein-dedicated analyses |
|------------|---------------------------|------|------|--------------------------|-----------------------------------|----------------------------------------|----------------------------------------|-----------------------------|
|            |                           |      |      |                          |                                   | Distance (nt)                        | Nearest reference MES scores          | Potential local splice effect        | Local MES scores |          | AGVGD | SIFT | MAPP | PolyPhen-2 | MutationTaster |
|            |                           |      |      |                          |                                   | Type                  |                                      | Δ (%)       | Δ (%)       | WT Var | WT Var | WT Var | WT Var |          |            |
| Patient_ID |                            |      |      |                          |                                   |                       |                                      |            |            |        |        |        |        |          |            |
| chr22      | 247                        | 458  | 5     | c.328                     | A>G                               | -84                    | 5'                                    | 10.1        | 10.1       | 90.6   | 90.6   | 0      | 1.2    | 1.50425  | 4.7         | 21.1          |
| chr1       | 46                         | 835  | 4     | c.35                      | A>G                               | -58                    | 5'                                    | 95          | 95         | 83.7   | 83.7   | 0      | -      | 16       | 27          | 77           |
| chr22      | 247                        | 511  | 217   | c.23                      | A>G                               | -33                   | 3'                                    | 7.5         | 7.5        | 93.6   | 93.6   | 0      | -      | -1.61    | 0.23        | -90           |
| chr1       | 46                         | 268  | 3'     | c.125                     | C>T                               | -12                   | 3'                                    | 89          | 89         | 87.5   | 87.5   | 0      | -      | 0.5      | 51          | 72           |
| chr1       | 459                        | 262  | 3'     | c.325                     | A>G                               | -10                   | 3'                                    | 70          | 70         | 90.1   | 90.1   | 0      | -      | -0.40     | 237         | -28           |

In order to predict the biological impact of the 25 VUS, RNA splicing- and protein-dedicated bioinformatics analyses were performed as described under Materials and Methods. The stars indicate exons that could not be tested in our minigene assay, either because of their terminal position (*, 1st or last exons) or because of their large size (**). Results shown in bold were considered as predictive of a potential variant-induced negative biological effect. MES, MaxEntScan; SSFL, Splice Site Finder-Like; nt, nucleotide; 3' or 5' ss, 3' splice site or 5' splice site; ESR, exonic splicing regulators; AGVGD, align-GVGD (C0, C15, C25, C35, C45, C55, or C65 with C65 most likely to interfere with function and C0 least likely), SIFT, Sorting Intolerant From Tolerant (tolerated or deleterious), MAPP, Multivariate Analysis of Protein Polymorphism (good or bad), PolyPhen-2, Polymorphism Phenotyping v2 (benign, possibly damaging or probably damaging), MutationTaster (polymorphism or disease causing), CRC: colorectal cancer.

In order to predict the biological impact of the 25 VUS, RNA splicing- and protein-dedicated bioinformatics analyses were performed as described under Materials and Methods. The stars indicate exons that could not be tested in our minigene assay, either because of their terminal position (*, 1st or last exons) or because of their large size (**). Results shown in bold were considered as predictive of a potential variant-induced negative biological effect. MES, MaxEntScan; SSFL, Splice Site Finder-Like; nt, nucleotide; 3' or 5' ss, 3' splice site or 5' splice site; ESR, exonic splicing regulators; AGVGD, align-GVGD (C0, C15, C25, C35, C45, C55, or C65 with C65 most likely to interfere with function and C0 least likely), SIFT, Sorting Intolerant From Tolerant (tolerated or deleterious), MAPP, Multivariate Analysis of Protein Polymorphism (good or bad), PolyPhen-2, Polymorphism Phenotyping v2 (benign, possibly damaging or probably damaging), MutationTaster (polymorphism or disease causing), CRC: colorectal cancer.

In order to predict the biological impact of the 25 VUS, RNA splicing- and protein-dedicated bioinformatics analyses were performed as described under Materials and Methods. The stars indicate exons that could not be tested in our minigene assay, either because of their terminal position (*, 1st or last exons) or because of their large size (**). Results shown in bold were considered as predictive of a potential variant-induced negative biological effect. MES, MaxEntScan; SSFL, Splice Site Finder-Like; nt, nucleotide; 3' or 5' ss, 3' splice site or 5' splice site; ESR, exonic splicing regulators; AGVGD, align-GVGD (C0, C15, C25, C35, C45, C55, or C65 with C65 most likely to interfere with function and C0 least likely), SIFT, Sorting Intolerant From Tolerant (tolerated or deleterious), MAPP, Multivariate Analysis of Protein Polymorphism (good or bad), PolyPhen-2, Polymorphism Phenotyping v2 (benign, possibly damaging or probably damaging), MutationTaster (polymorphism or disease causing), CRC: colorectal cancer.

In order to predict their biological impact, RNA splicing- and protein-dedicated bioinformatics analyses were performed as described under Materials and Methods. Results shown in bold were considered as predictive of a potential variant-induced negative biological effect. MES, MaxEntScan; SSFL, Splice Site Finder-Like; nt, nucleotide; 3' or 5' ss, 3' splice site or 5' splice site; ESR, exonic splicing regulators; AGVGD, align-GVGD (C0, C15, C25, C35, C45, C55, or C65 with C65 most likely to interfere with function and C0 least likely), SIFT, Sorting Intolerant From Tolerant (tolerated or deleterious), MAPP, Multivariate Analysis of Protein Polymorphism (good or bad), PolyPhen-2, Polymorphism Phenotyping v2 (benign, possibly damaging or probably damaging), MutationTaster (polymorphism or disease causing), CRC: colorectal cancer.

In order to predict their biological impact, RNA splicing- and protein-dedicated bioinformatics analyses were performed as described under Materials and Methods. Results shown in bold were considered as predictive of a potential variant-induced negative biological effect. MES, MaxEntScan; SSFL, Splice Site Finder-Like; nt, nucleotide; 3' or 5' ss, 3' splice site or 5' splice site; ESR, exonic splicing regulators; AGVGD, align-GVGD (C0, C15, C25, C35, C45, C55, or C65 with C65 most likely to interfere with function and C0 least likely), SIFT, Sorting Intolerant From Tolerant (tolerated or deleterious), MAPP, Multivariate Analysis of Protein Polymorphism (good or bad), PolyPhen-2, Polymorphism Phenotyping v2 (benign, possibly damaging or probably damaging), MutationTaster (polymorphism or disease causing), CRC: colorectal cancer.

In order to predict their biological impact, RNA splicing- and protein-dedicated bioinformatics analyses were performed as described under Materials and Methods. Results shown in bold were considered as predictive of a potential variant-induced negative biological effect. MES, MaxEntScan; SSFL, Splice Site Finder-Like; nt, nucleotide; 3' or 5' ss, 3' splice site or 5' splice site; ESR, exonic splicing regulators; AGVGD, align-GVGD (C0, C15, C25, C35, C45, C55, or C65 with C65 most likely to interfere with function and C0 least likely), SIFT, Sorting Intolerant From Tolerant (tolerated or deleterious), MAPP, Multivariate Analysis of Protein Polymorphism (good or bad), PolyPhen-2, Polymorphism Phenotyping v2 (benign, possibly damaging or probably damaging), MutationTaster (polymorphism or disease causing), CRC: colorectal cancer.
within these families [23, 54, 55]. Co-segregation analyses may add clues in our understanding whether this germline variant is implicated in CRC predisposition. Finally, we did not find pathogenic variants in \( \text{POLE} \) in our cohort, which is in contrast to what has been described in families with high burden of CRC adenomas and carcinomas in addition to extra-colonic cancers [56].

According to the Prospective LS Database (PLSDB), a total of 125 Norwegian families had a demonstrated pathogenic variant in either \( \text{MLH1} \) (\( n = 21 \)), \( \text{MSH2} \) (\( n = 52 \)), \( \text{MSH6} \) (\( n = 36 \)), or \( \text{PMS2} \) (\( n = 16 \)) [25]. On the other hand, a large portion of high-risk CRC families without pathogenic variant in MMR or \( \text{EPCAM} \) genes may be explained by a polygenic model involving a combination of multiple genomic risk factors, including the effect of either low-penetrance susceptibility alleles [57], high-penetrance genes which have not been tested, or the effect of environmental factors. In addition, emerging data suggest that CRC cases negative for pathogenic MMR variants may contain a significantly higher number of copy-neutral loss of heterozygosity (cnLOH) regions, some located within well-known oncogenes and tumor suppressor genes, compared to cases of sporadic CRC [58]. These genomic variations, which were not investigated in this study, may provide an additional explanation for high-risk CRC phenotypes without MMR or \( \text{EPCAM} \) pathogenic variants.

Recent NGS studies described the presence of heterozygous pathogenic \( \text{BRCA1/2} \) or \( \text{APC} \) variants as well as biallelic \( \text{MUTYH} \) alterations in individuals with clinical features resembling those of LS [5, 22]. More precisely, those studies reported that 7% of patients with CRC carried pathogenic variants in non-LS genes, including 1.0% with \( \text{BRCA1/2} \) mutations, and nearly two thirds of probands with high-penetrance non-LS mutations lacked clinical histories suggestive of their respective syndromes [5].

From 34 high-risk CRC individuals, our NGS panel testing identified one patient that carried a pathogenic variant in a gene with reportedly moderate penetrance. Our finding is in line with the mutation frequency (6%) in non-LS cancer susceptibility genes for individuals undergoing LS genetic testing [21] and 4% of patients with BC tested negative for \( \text{BRCA1/2} \) genes [23]. Our results may have implications for an appropriate genetic
counseling and follow-up of the patients and family members.

Besides the likely pathogenic CHEK2 variant, we identified a total of 25 variants in our cohort for which there were not so much data as to their clinical significance. We thus undertook bioinformatics analyses in an attempt to predict the biological impact of these Class 3 variants, both at the RNA and protein level, the ultimate goals being: (i) to discriminate pathogenic from non-pathogenic alterations in this set of variants and (ii) to further pinpoint the genetic determinants of high risk CRC in our cohort. On one hand, our RNA splicing-dedicated bioinformatics evaluation predicted that 2 out of the 25 VUS identified in this study (NOTCH3 c.5854G > A, p.V1952 M and MAP3K1 c.764A > G, p.N255S) could potentially affect RNA splicing. These two variants were then experimentally analyzed by performing minigene splicing assay. Our results revealed that neither variant altered the splicing pattern of the representative minigenes, suggesting that they do not affect the splicing of NOTCH3 or MAP3K1 transcripts. Additional experiments based on the analysis of RNA from carriers of these variants will be important to verify our minigene results. On the other hand, our protein-dedicated bioinformatics analysis yielded 8 consistent predictions (2 VUS predicted as deleterious and 6 as benign) and several conflicting results that were not explored further.

In this scenario, not only functional tests, but also co-segregation studies will be key to understanding whether the VUS detected in this work are non-pathogenic or otherwise have a causal or a modifying effect. Importantly, we do not yet have evidence-based guidelines for the majority of the genes carrying the VUS identified in this study and, in more general terms, the impact of low- to moderate-penetrance pathogenic variants with respect to clinical management is not fully understood. Most of these variants may in the future be reclassified as deleterious or benign, but in the meantime, they cannot be used to make clinical decisions [59]. Informed (re)classification of VUS in cancer-associated genes may cater to more appropriate risk-management, and may provide significant clues for the identification of additional patients carrying such uncommon variants.

NGS panel testing may benefit patients with a personal or family history compatible with more than one recognized CRC inherited syndrome. The CRC risk management strategy for these individuals is not yet available and there is a need to identify new high-, moderate-, and low- penetrance gene variants that may affect the risk of CRC or LS-associated tumors in non-MMR pathogenic carriers. The identification of such gene variants in combination with family history may contribute to more intensive surveillance and improved prevention [23].

Conclusions

Our study provides information on genetic locus that might possibly be related to cancer susceptibility, demonstrating that genes presently not routinely tested may be important for capturing cancer predisposition in these patients. In addition, we stratified 25 VUS by the use of RNA splicing- and protein-dedicated in silico analyses. Further studies are necessary for making reliable estimates of cancer risk for the VUS found in this study and allowing appropriate genetic counseling for the patients and their relatives.

Surveillance for early cancer detection is essential to ensure optimal survival for patients afflicted with familial cancers. Our findings pinpoint the need of more studies to unravel the mechanisms underlying the development of CRC in high-risk patients and the identifying for new cancer predisposition genes.

Additional file

**Table S1.** Primers used in the pCAS2 minigene splicing assay. (DOCX 15 kb)

**Abbreviations**

ACMG: American college of medical genetics and genomics; AMS: Amsterdam criteria; BC: Breast cancer; CRC: Colorectal cancer; IHC: Immunohistochemistry; LS: Lynch syndrome; MAF: Minor allele frequency; MMR: Mismatch repair genes; MSI: Microsatellite instability; NGS: Next generation sequencing; PLSDB: Prospective lynch syndrome database; SNV: Single nucleotide variant; VUS: Variants of unclassified significance

**Acknowledgements**

We thank the families for their participation and contribution to this study.

**Funding**

This work was supported by the Radium Hospital Foundation (Oslo, Norway), Helse Sør-Øst (Norway), the French Association Recherche contre le Cancer (ARC), the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (Gefluc), the Association Nationale de la Recherche et de la Technologie (ANRT, CIFRE PhD fellowship to H.T.) and by the OpenHealth Institute.

**Availability of data and materials**

All data generated or analyzed during this study are included in the manuscript.

**Authors’ contributions**

All authors have taken part in the different steps of the study: MDV, PM and EH designed the study, AM, HT performed in silico splicing predictions and the minigene assays, POE performed validation experiments, MM, AN and EHF performed in silico protein predictions, SN, DV performed the sequence analysis. MDV drafted the manuscript and all have read, revised and approved the manuscript.

**Ethics approval and consent to participate**

Ethical approval for the study was granted by the Norwegian Data Inspectorate and Ethical Review Board (ref 2015/2382). All examined patients signed an informed consent for their participation in the study.
Consent for publication
Not Applicable.

Competing interests
The authors declare that they have no competing interests

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. 2Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. 3Interactive Biotechnology Centre, Faculty of Medicine, Rouen, France. 4Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Munich, Germany. 5Department of Human Medicine, Université Witten, Herdecke, Germany. 6Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. 7Department of Informatics, University of Oslo, Oslo, Norway. 8Department of Informatics, University of Oslo, Oslo, Norway. 9Inserm-U1245, UNIROUEN, Normandie Univ, Rouen, Normandy Centre for Genomic and Personalized Medicine, Rouen, France. 10Department of Informatics, University of Oslo, Oslo, Norway. 8Department of Informatics, University of Oslo, Oslo, Norway.

Received: 30 June 2017 Accepted: 24 January 2018
Published online: 20 February 2018

References
1. Llor X, Pons E, Xicola RM, Castells A, Alenda C, Pinol V, Andreu M, Castellvi-Bel S, Paya A, Jover R, et al. Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway. Clin Cancer Res. 2005;11(20):7304–10.
2. Balmana J, Castells A, Cervantes A, Group EGW. Familial colorectal cancer risk: ESMO Clinical Practice Guidelines. Ann Oncol. 2010;21(Suppl 5):v78–81.
3. Buchanan DD, Tan YY, Walsh MD, Glendening M, Metcalfe AM, Ferguson K, Arnold ST, Thompson BA, Lose FA, Parsons MT, et al. Tumor mismatch repair Immunohistochemistry and DNA MLH1 Methylation testing of patients with endometrial cancer diagnosed at age younger than 60 years optimizes trage for population-level Germline mismatch repair gene mutation testing. J Clin Oncol. 2014;32(2):90.
4. Patel SG, Ahnen DJ. Familial colon cancer syndromes: an update of a rapidly evolving field. Curr Gastroenterol Rep. 2012;14(5):428–38.
5. Yurgelun MB, Kulkhe MK, Fuchs CS, Allen BA, Ulu H, Horsnick J, Ukegbu CI, Braas LK, McNamara PM, Mayer RJ, et al. Cancer susceptibility gene mutations in individuals with colorectal cancer. J Clin Oncol. 2017;35(10):JCO2016710012.
6. Kasper RP, Vissers LE, Venkatachalam R, Bodmer D, Hoenselaar E, Goossens M, Haufe A, Kamping E, Niessen RC, Hogervorst FB, et al. Recurrence and variability of germline EPCAM deletions in Lynch syndrome. Hum Mutat. 2011;32(4):407–14.
7. Vassen HF, Mecklin JP, Khan PM, Lynch HT. The international collaborative group on hereditary non-polyposis colorectal cancer (IHC-NHPPC). Dis Colon Rectum. 1991;34(5):424–5.
8. Moller P, Evans G, Haines N, Vassen H, Reis MM, Anderson E, Apld J, Hodgson S, Eccles D, Olsson H, et al. Guidelines for follow-up of women at high risk for inherited breast cancer: consensus statement from the biomed 2 demonstration Programme on inherited breast cancer. Dis Markers. 1999;15(1–3):207–11.
9. Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Peruchno M, Smyrk T, Sobin L, et al. A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst. 1997;89(23):1758–62.
10. Ulmar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (lynch syndrome) and microsatellite instability. J Natl Cancer Inst. 2004;96(4):261–8.
11. Kobayashi OH, Ohno S, Sasaki Y, Matsuura M. Hereditary breast and ovarian cancer susceptibility genes (review). Oncol Rep. 2013;30(3):1019–29.
12. Lindor NM. Familial colorectal cancer type X: the other half of hereditary nonpolyposis colon cancer syndrome. Surg Oncol Clin N Am. 2005;14(4):637–.
13. Rodríguez-Soler M, Perez-Carbonell L, Guarinos C, Zapater P, Castillo A, Barbera VM, Juarez M, Bessa X, Xicola RM, Clofent J, et al. Risk of cancer in cases of suspected Lynch syndrome without Germ-line mutation. Gastroenterology. 2013;144(5):S26.
14. Dominguez-Valentin M, Therkildsen C, Da Silva S, Nilbert M. Familial colorectal cancer type X: genetic profiles and phenotypic features. Mod Pathol. 2015;28(1):30–6.
15. da Silva FC, Ferreira JRD, Torezan GT, Figueiredo MCP, Santos EMM, Nakagawa WT, Brianece RC, de Oliveira LP, Begnani MD, Aguiar S, et al. Clinical and molecular characterization of Brazilian patients suspected to have Lynch syndrome. PLoS One. 2015;10(10):e013973.
16. Amendola LM, Janvik GP, Leo MC, Maclachlin HM, Akkari Y, Amrall MD, Berg JS, Biswas S, Bowling WM, Conlin UK, et al. Performance of ACMG-AMP variant-interpretation guidelines among nine Laboratories in the Clinical Sequencing Exploratory Research Consortium (vol 98, pg 1067, 2016). Am J Hum Genet. 2016;99(2):1247.
17. Pearlman R, Frankel WL, Swanson B, Zhao W, Yilmaz A, Miller K, Bacher J, Bigley C, Nielsen L, Goodfellow PJ, et al. Prevalence and Spectrum of Germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. JAMA Oncol. 2016;3(4):646–7. https://doi.org/10.1001/jamaoncol.2016.5194.
18. Picelli S, Lorenzo Bermejo J, Chang-Claude J, Hofmeister M, Fernandez-Rozadilla C, Carraoedo A, Castells A, Castelli-Bel S, Memesbers of EC- GCOTSGA, Naccarati A, et al. Meta-analysis of mismatch repair polymorphisms within the cognet consortiom for colorectal cancer susceptibility. PLoS One. 2015;10(8):e013897.
19. Ballestre V, Boardman L. Next generation multigene panel testing: the next step for identification of hereditary colorectal cancer syndromes? Gastroenterology. 2015;149(3):S26–8.
20. Kurian AW, Hare EE, Mills MA, Kingham KE, McPherson L, Whittemore AS, McGivney V, Ladabaum U, Kebayashi Y, Lincoln SE, et al. Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. J Clin Oncol. 2014;32(19):2001–8.
21. Yurgelun MB, Masciari S, Joshi VA, Mercado RC, Lindor NM, Gallinger S, Hopper JL, Jenkins MA, Buchanan DD, Newcomb PA, et al. Germline TP53 mutations in patients with early-onset colorectal cancer in the colon cancer family registry. JAMA Oncol. 2015;12(12):214–21.
22. Yurgelun MB, Allen B, Kaldate RR, Bowles KR, Judkins T, Kauhshik P, Roa BB, Wenstrup RJ, Hartman AR, Syngal S. Identification of a variety of mutations in cancer predisposition genes in patients with suspected Lynch syndrome. Gastroenterology. 2015;149(3):e604–13. e620
23. Tung N, Lin NJ, Kidd J, Allen BA, Singh N, Wenstrup RJ, Hartman AR, Winer EP, Garber JE. Frequency of Germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. J Clin Oncol. 2016;34(13):1460–461.
24. Moller P, Seppala T, Bernstein I, Holinski-Feder E, Sala P, Evans DG, Lindblom A, Macrae F, Blanco I, Sijmons R, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. Gut. 2015;64(3):464–72.
25. Moller P, Seppala T, Bernstein I, Holinski-Feder E, Sala P, Evans DG, Lindblom A, Macrae F, Blanco I, Sijmons R, et al. Incidence of and survival after subsequent cancers in carriers of pathogenic MMR variants with previous cancer: a report from the prospective Lynch syndrome database. Gut. 2017;66(1):1657–64. https://doi.org/10.1136/gutjnl-2016-311403.
26. Moller P, Seppala TT, Bernstein I, Holinski-Feder E, Sala P, Garrett EV, Lindblom A, Macrae F, Blanco I, Sijmons RH, et al. Cancer risk and survival in path-MMR carriers by gene and gender up to 75 years of age: a report from the prospective Lynch syndrome database. Gut. 2017; https://doi.org/10.1136/gutjnl-2017-314057. [Epub ahead of print]
27. Lynch HT, Riley DB, Weissman SM, Coronel DD, Kinarcky S, Lynch JF, Shaw TG, Rubinstein WS. Hereditary nonpolyposis colorectal carcinoma (HNPCC) and HNPCC-like families: problems in diagnosis, surveillance, and management. Cancer Am Cancer Soc. 2004;100(1):53–64.
28. Dominguez-Valentin M, Nakken S, Tudeb H, Vodak D, Ekström PO, Nissen AM, Morak M, Holinski-Feder E, Martin M, Moller P, et al. Potentially pathogenic germline CHEK2 c.319+2T>C among many early-onset colorectal cancer families. Familial Cancer. 2018;17(1):141–53. https://doi.org/10.1007/s10690-017-0111-0.
29. Dominguez-Valentin M, Evans DGR, Nakken S, Tudeb H, Vodak D, Ekström PO, Nissen AM, Morak M, Holinski-Feder E, Martin M, Moller P, Hovig E. Genetic variants of prospectively demonstrated phenocopies in BRCA1/2
