PMS2-associated Lynch syndrome : the odd one out

Broeke, S.W. ten


3

RISK MODIFYING FACTORS

3.1

SNP association study in PMS2-associated Lynch syndrome

Familial Cancer, 2017

Sanne W. ten Broeke, Fadwa A. Elsayed, Lisa Pagan, Maran J.W. Olderode-Berends, Encarna Gomez Garcia, Hans J.P. Gille, Liselot P. van Hest, Tom G.W. Letteboer, Lizet E. van der Kolk , Arjen R. Mensenkamp, Theo A. van Os, Liesbeth Spruijt, Bert J.W. Redeker, Manon Suerink, Yvonne J. Vos, Anja Wagner, Juul T. Wijnen, E.W. Steyerberg, Carli M.J. Tops, Tom van Wezel, Maartje Nielsen

ABSTRACT

Objective

Lynch syndrome (LS) patients are at high risk of developing colorectal cancer (CRC). Phenotypic variability might in part be explained by common susceptibility loci identified in Genome Wide Association Studies (GWAS). Previous studies focused mostly on *MLH1*, *MSH2* and *MSH6* carriers, with conflicting results. We aimed to determine the role of GWAS SNPs in *PMS2* mutation carriers.

Methods

A cohort study was performed in 507 *PMS2* carriers (124 CRC cases), genotyped for 24 GWAS SNPs, including SNPs at 11q23.1 and 8q23.3. Hazard ratios (HRs) were calculated using a weighted Cox regression analysis to correct for ascertainment bias. Discrimination was assessed with a concordance statistic in a bootstrap cross-validation procedure.

Results

Individual SNPs only had non-significant associations with CRC occurrence with HRs lower than 2, although male carriers of allele A at rs1321311 (6p21.31) may have increased risk of CRC (HR=2.1, 95%CI: 1.2-3.0). A polygenic risk score (PRS) based on 24 HRs had an HR of 2.6 (95%CI:1.5-4.6) for the highest compared to the lowest quartile, but had no discriminative ability (c statistic 0.52).

Conclusion

Previously suggested SNPs do not modify CRC risk in *PMS2* carriers. Future large studies are needed for improved risk stratification among Lynch syndrome patients.

INTRODUCTION

Lynch syndrome (LS) accounts for 2-4% of all CRCs and is characterized by a high risk for developing malignancies, most notably colorectal cancer (CRC) and endometrial cancer (EC). The underlying cause is a germline mutation in one of the mismatch repair (MMR) genes: *MLH1*, *MSH2* (*EPCAM*), *MSH6* or *PMS2*. Mutations in all MMR genes are associated with a significantly increased cancer risk compared to the general population, although *MSH6* and *PMS2* carriers show lower penetrance compared to *MLH1* and *MSH2* carriers.[1,3] Within and between family variability is commonly observed and a range of theories have been proposed to explain the phenomenon, such as genotype-phenotype correlations, parent-of-origin effects, lifestyle factors and the influence of common susceptibility loci. The latter, mainly single nucleotide polymorphisms (SNPs), were identified in genome wide association studies (GWAS) in large cohorts consisting of sporadic CRC cases.[4] Among these candidate SNPs, previous studies have identified statistically significant effects of multiple SNPs in LS patients, and independent studies replicated the effect of SNPs rs3802842 (11q23.1) and rs16892766 (8q23.3) among *MLH1* carriers.[5, 6] It should be noted, however, that others have failed to replicate these findings.[7, 8] Although the latter studies analyzed cohorts of similar size to our own, few or no *PMS2* carriers were included.[7] Due to a relatively low penetrance and high phenotypic variability, this specific subset of LS patients might be of particular interest.[9] In a previous study among 377 *PMS2* carriers, we found age at CRC diagnosis to vary widely (range 26-86 years) and mean age of index carriers and mutation-positive family members differed by 10 years.[3] In the current study, we aim to determine whether these SNPs modify CRC risk in a large cohort of *PMS2* mutation carriers.

MATERIAL AND METHODS

Sample collection

DNA extracted from leucocyte DNA was collected from 8 Dutch family cancer clinics. Index carriers included in this study were sent in between 2007 and 2016 to the Clinical Genetics department, because of a clinical suspicion of LS, e.g. LS-associated cancer at a young age and/or a positive family history. Mutation analysis was initiated based on the presence of histological hallmarks (microsatellite instability (MSI) or loss of PMS2 expression in the tumor) and/or when the family complied with the Bethesda Criteria.[10] Participating clinics provided DNA samples and clinical data on CRC, age at diagnosis,

other cancer development and polypectomy. Controls were defined as carriers that were tested pre-symptomatically, after a pathogenic mutation was identified in the index carrier of the family. All carriers are referred to gastroenterology departments after the diagnosis has been established, which then adhere to international surveillance guidelines, i.e. colonoscopies every 1-2 years from 25 years of age.[11] Data was analyzed anonymously. The study was approved by the medical ethical committee of Leiden University Medical Centre, protocol ID P01-019.

Genotyping

PMS2 genotyping in this cohort was carried out as previously described.[3] SNP genotyping was done at the LUMC laboratory using a KASp genotyping assay (LGC Genomics, Hoddesdon, UK). Primers were designed using Primerpicker (KBioscience, Hoddesdon, UK) and are available upon request. All oligonucleotides were obtained from Eurofins Genomics (Ebersberg, Germany). Genotypes were called using the CFX manager software v3.0 (Bio-Rad, Veenendaal, the Netherlands).

Statistical analysis

PMS2 carriers were analyzed as a birth cohort. A Cox-proportional hazards regression model was fitted to estimate hazard ratios (HRs), with age at CRC as endpoint and SNP genotype as independent variable. Patients without CRC were censored at the last age known to be alive. The HR was calculated separately for heterozygous and homozygous carriers of the risk allele, with homozygotes of the non-risk allele as reference category. We also calculated the per allele HR (additive model). Cox-regression analyses were also stratified for gender. These sub-analyses only includes a per allele (additive) model, due to multiple testing risks. Missing age at CRC diagnosis (n=3) was imputed using median age of CRC in the general population (age 70, n=2) or set at one year before death (n=1). The proportional hazards assumption was investigated by examining the scaled Schoenfeld residuals with a formal statistical test and by visual inspection.

Previous studies have described the oversampling of cases in clinic-based cohorts. Moreover, affected family members are more likely to be tested for the mutation and this too results in oversampling of cases. To adjust for this non-random sampling, we used a weighted cohort approach as previously described.[12] Standard errors were corrected for familial clustering of risk by using the Huber-White sandwich estimator.[13] We also calculated two polygenic risk scores (PRS)[14] based on 1) the odds ratios (ORs) reported in the meta-analysis by Ma et al (Supplemental Table 1) and 2) based on our HRs from the current study.[4] Kaplan Meier (KM) and Cox regression analysis

were concurrently performed. A concordance statistic was calculated to assess the discriminative value of the 24 SNP model. The optimism in the concordance statistic was estimated by fitting the model in each of 500 bootstrap samples (drawn with replacement), and validation in the original sample. Analyses were initially performed for patients with complete data. We also performed imputation of missing values based on the correlation structure between SNPs and with the outcome (transcan function in R software, version 3.2). Since results were similar, we only present complete case results.

Lastly, a post hoc power analysis was performed to assess the chance of finding significantly increased risks using the collected cohort, which contains all currently known *PMS2* mutation carriers in the participating centers (Supplemental Figure 1). We had at least 80% power to find an HR of 1.5 for the majority of SNPs. A more detailed description of the statistical analysis is available in the supplementary methods (Supplemental File 1). Statistical analyses were performed using STATA version 14 (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP) and R software (version 3.2, using the rms library).

RESULTS

Cohort

In total, 521 samples from carriers with a germline *PMS2* mutation were genotyped, derived from 152 families. Fourteen patients were excluded because 1) they were younger than 25 years at censoring and were therefore not yet at risk of developing CRC (n=11) or 2) insufficient clinical data was available (n=3, including one CRC case). The analyzed cohort consisted of 124 cases (*PMS2* carriers with CRC) and 383 controls (*PMS2* carriers without CRC), with attributed person years of 6527 and 19549, respectively. Person-years were calculated until age of CRC for cases (*PMS2* carriers with CRC, n=125), and age at polypectomy, age of death, or last known age alive (whichever occurred first) for controls (*PMS2* carriers without CRC, n=1, n=1 and n=381 respectively). The mean age was 52.5 for CRC cases and 51.0 for non-cases (Table 1). For a detailed description of the families including genotypes see Supplemental Tables 2a+2b.

TABLE 1 Cohort description

	No CRC – controls (n=383)	CRC - cases (n=124)	All (n=507)
Sex			
Male	133 (34%)	60 (48%)	193 (38%)
Female	250 (65%)	64 (52%)	314 (62%)
Age (CRC or censoring)			
Mean (s.d.)	51.0 (14.2)	52.5 (12.7)	
Range	25-88	27-88	
Index carrier			
Yes	38 (10%)	89 (72%)	127 (25%)
No, family member	345 (90%)	35 (28%)	380 (75%)
Other cancers (no. of carriers)			
Endometrial cancer#	30	9	39
Ovarian#	4	0	4
Duodenal cancer#	4	2	6
Breast#	10	3	13
Urothelial#	4	2	6
Esophagus	1	0	1
Leukemia	0	3	3
Testis	2	0	2
Prostate	1	1	2
Vagina	0	1	1
Mesothelioma	0	1	1

'Index carrier' means the first person to be tested. Incidence of cancer in the group of index carriers without CRC: 20 endometrial cancers, 4 ovarian cancers, 3 breast cancers, 3 cancers of the small intestine, 1 testis cancer and 1 carcinoid. Ten of these index carriers had not developed any cancer at the time of DNA diagnostics; they were tested because of polyps at an early age or because they had an (affected) deceased family member. # Lynch syndrome-associated tumor.

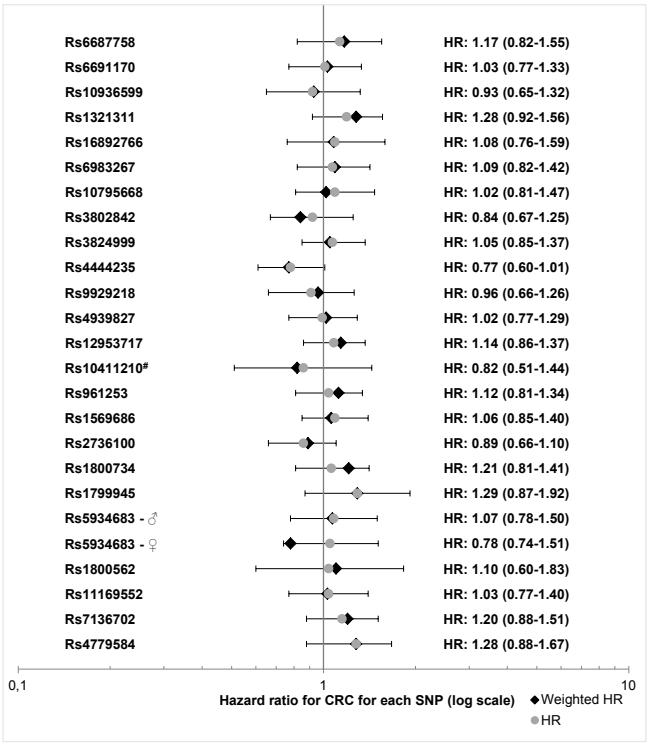


FIGURE 1 Forest plot of HRs for all SNPs. Note: rs5934683 lies on the X chromosome and was therefore stratified for gender. *SNPs previously associated with increased risk in MLH1 mutation carriers. #Reference category: homozygous for risk allele (due to low number of homozygous carriers of the non-risk allele). HR: Hazard Ratio.

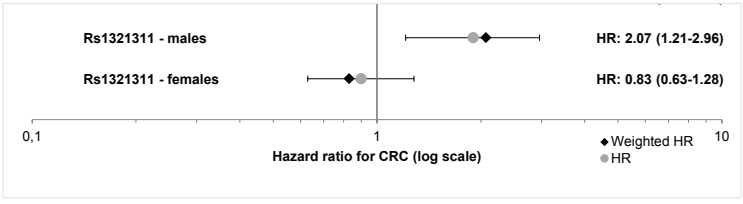


FIGURE 2 Forest plot of HRs for rs1321311. Note: p=.005 for males. HR: Hazard Ratio.

Hardy-Weinberg equilibrium

Two SNPs, rs1048943 (15q24.1) and rs4925386 (20q13), were not in Hardy Weinberg equilibrium (HWE). Violation of the HWE was present in both cases and controls and as this might be the result of a genotyping error, these SNPs were removed from the analysis. Ultimately, 24 SNPs were included in the final analysis.

Risk of colorectal cancer

None of the SNPs individually showed a clear risk modifying effect (Figure 1, Supplemental Table 1). There was a difference in HR between male and female *PMS2* carriers for rs1321311 (6p21.31), with an HR for the each additional A allele of 2.07 (95%CI: 1.21-2.96, p=0.005) and 0.83 (95%CI:0.63-1.28, p=0.56) for males and females, respectively (Figure 2: Forest plot, Supplemental Figure 2: KM curve).

Combination of rs3802842 & rs16892766

A previous meta-analysis reported a significant pairwise effect on CRC risk of rs3802842 (11q23.1) and rs16892766 (8q23.3) in *MLH1* mutation carriers.[6] The HR in the additive model for this combination in our *PMS2* cohort was 0.95 (95%CI:0.80-1.25, p=0.99). For carriers of more than three risk alleles the HR was 1.58 (95%CI: 0.55-3.39) compared to patients with no risk alleles, see Figure 3 for a comparison of previously publishes HRs and results from this study. The mean age at CRC diagnosis for 0, 1, and more than 1 risk alleles was 52.8; 52.9 and 50.4 respectively. The corresponding median ages were 54 (interquartile range (IQR): 43-62), 51 (IQR:43-63) and 47 (IQR:39-63). There was a statistically non-significant difference between the median age of CRC diagnosis between male and female carriers of two or more risk alleles, namely 53 (IQR:39-64, n=7) and 43.5 years to age (IQR:38-63, n=10, p=0.56, Mann-Whitney test).

Polygenic risk score

The polygenic risk score was calculated for 444 *PMS2* carriers with complete genotyping. The medians for PRS1 (meta-analysis derived ORs) were -0.12 (interquartile range (IQR):-0.48-0.30) for controls and -0.03 (IQR:-0.39-0.40) for CRC cases. The HRs for group 2 (second and third quartile) and 3 (fourth quartile) were 1.33 (95%CI:0.76-2.33) and 1.50 (95%CI:0.82-2.72) respectively (Table 2). The medians for PRS2 (based on HRs from our own data) were 0.30 (IQR:-0.057-0.55) for controls and 0.51 (IQR:0.068-0.75) for CRC cases. The corresponding HRs for group 2 and 3 were 1.05 (95%CI:0.59-1.89) and 2.62 (95%CI:1.49-4.60) respectively (Table 2). The KM curves for PRS1 and 2 are shown in and Supplemental Figure 3 and Figure 4, respectively. The difference between survival curves was highly significant for PRS2 (p<0.0001). The optimism-corrected c statistic was only 0.52, indicating no discriminatory value.

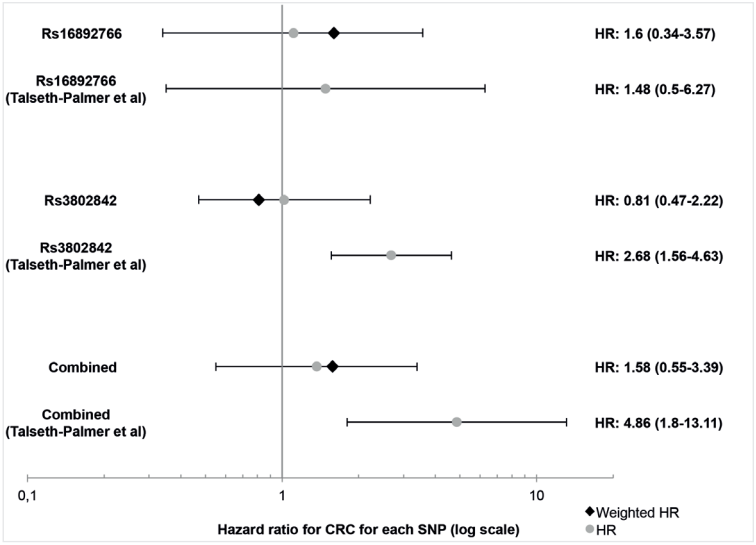


FIGURE 3 Forest plot of HRs for rs3802842 and rs16892766. Note: For the combination of the two SNPs, the plotted HR represents a comparison for carriers of three vs. no risk alleles. HR: Hazard Ratio.

TABLE 2 Polygenic risk scores

	PRS category	Controls	CRC	Expected events	HR	p for HR
PRS1	Median (IQR)	-0.12 (-0.48-0.30)	-0.03 (-0.39-0.40)			
	1st quartile	84	22	28	ref	0.41
	2d & 3d quartile	167	54	52	1.33 (0.76-2.33)	0.31
	4th quartile	84	33	28	1.50 (0.82-2.72)	0.19
PRS2	Median (IQR)	0.30 (-0.057-0.55)	0.51 (0.068-0.75)			
	1st quartile	84	18	26	ref	<0.0001
	2d & 3d quartile	165	38	53	1.05 (0.59-1.89)	0.86
	4th quartile	84	53	30	2.62 (1.49-4.60)	0.001

PRS: polygenic risk score. IQR: interquartile range.
PRS1: Weighted on odds ratios from general population, i.e. in sporadic CRC cases.
PRS2: Weighted on hazard ratios from this study
Log rank survival curves PRS1: p=0.32
Log rank survival curves PRS2: p<0.0001

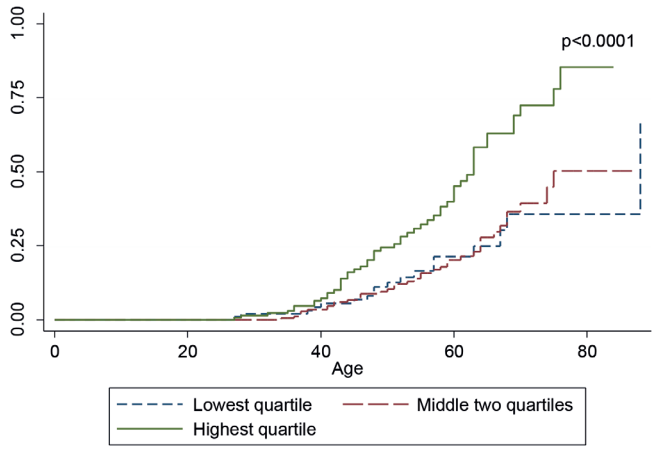


FIGURE 4 Kaplan Meier survival curve with endpoint colorectal cancer for PRS2 . Note: This plot compares curves for the lowest, the two middle and the highest quartile of the PRS. PRS2 is based on hazard ratios from the current study. HR: Hazard Ratio. PRS: polygenic risk score.

DISCUSSION

PMS2 carriers currently represent a relatively small proportion of LS patients. However, the number of *PMS2*-associated LS cases is expected to rise with the implementation of population-based screening protocols for all CRC below age 70. Identification of *PMS2* carriers has been challenging in the past due to difficulties in mutation analysis, a milder phenotype and many families not fulfilling clinical selection criteria.[3, 9, 15-17] Obtaining a better understanding of the specific *PMS2*-associated phenotype is particularly relevant, as it appears to differ markedly from phenotypes associated with other MMR mutations. Unfortunately, we were unable to confirm any risk modifying effects of rs3802842 (11q23.1) and rs16892766 (8q23.1), two SNPs previously shown to be associated with enhanced risk in *MLH1* mutation carriers.[5, 6] Studies in *MLH1* mutation carriers reported that a higher number of risk alleles in a carrier is associated with a younger onset of disease (28 years younger for 3 compared to 0 risk alleles).[6] In our cohort, mean ages where 52.8 and 50.4 for 0 compared to more than 1 risk allele, respectively. As such, there seems to be no clinical utility of rs3802842 and rs16892766 in risk stratification for *PMS2* carriers.

Many studies on (genetic) modifiers in LS patients focus on *MLH1* and *MSH2*, or *MSH6* carriers, while *PMS2* is seldom analyzed. The only study to include *PMS2* carriers (n=40) found that carriers of the G-alleles of rs10795668 (10p14) and rs9929218 (16q22.1) were at lower risk of CRC, a notable finding in that this is the opposite effect compared to sporadic CRC.[7] The authors conceded that their results should be confirmed in larger studies. As these findings have not been confirmed in our much larger cohort, we suggest that these previous findings may indeed have been false positives due to the small number of carriers included.

A relevant question is why our study did not confirm reported findings of previous studies of MMR carriers. One explanation might be that although patients have germline mutations in genes with similar functions, carriers are affected by genetic modifiers in different ways. Indeed, comparable studies in *BRCA1* or *BRCA2* mutation carriers have resulted in the identification of SNPs that clearly modify breast cancer risk. However, *BRCA2* carriers appear unaffected by SNPs that confer an increased breast cancer risk in *BRCA1* carriers, even though both genes play a role in homologous recombination.[18] This could also hold for MMR mutation carriers, as illustrated by the observation that while rs3802842 and rs16892766 may increase risk in *MLH1* carriers, they do not appear to have an effect in *MSH2* or *PMS2* carriers.[5, 6] Researchers should therefore concentrate on building cohorts large enough to analyze Lynch patients in a gene-stratified manner.

Gender stratification in our cohort led to the notable finding that male carriers of allele A at locus rs1321311 (6p21.31) show a per allele HR of 2.07 (95%CI. 1.21-2.96), while the HR for females was 0.83 (95%CI: 0.63-1.28). This SNP has been linked to the *CDKN1A* gene that encodes the p21 protein. p21 is involved in several (p53-independent) pathways as a tumor suppressor, although it also has oncogenic characteristics.[19, 20]Interestingly, down-regulation of p21 is inversely associated with MSI, the hallmark of Lynch-associated tumors. One study found that a larger proportion of Lynch-associated CRCs expressed p21 compared to sporadic CRCs (80% vs 31%).[21] However, a recent study reported on expressive Quantitative Trait Loci (eQTL) in colonic tissue based on data from the GTEx project portal (http://www.gtexportal.org/home/) and did not find a statistically significant effect of rs1321311 on CDKN1A expression in sigmoid and transverse colon tissue (p=0.84 and p=1.00 respectively).[22, 23] It is also unclear why this effect only appears to be present in male *PMS2* carriers. Although no gender difference was noted by the meta-analysis that identified the SNP [19], it is possible that gender differences exist, as CRC risk in the general population and in Lynch patients is known to be higher in men compared to women.[24, 25] Indeed, mutations are more often found in males than females when assessed for Lynch syndrome.[26] Another possible explanation for this gender specific effect might lie in the effect of other risk modifiers. It is perceivable that other factors than SNPs have a stronger influence on (colorectal) cancer development in women, such as hormonal factors.[27, 28] Unfortunately data on hormone levels or other factors previously shown to modify cancer risk in LS such as medication use (e.g. aspirin) or environmental factors were unavailable for analysis and as such we were unable to correct for this.[29-33] Similarly, data on smoking and BMI were only available for a small proportion of carriers (n=131, 26%). It should be emphasized that all results after gender stratification should be interpreted with caution because of small sample size and multiple testing. This could have led to false associations. Further studies are needed to validate these findings.

We also investigated the effect of the 24 SNPs on CRC risk in the *PMS2* cohort by means of a polygenic risk score (PRS). While there did not appear to be a significant effect of the PRS based on ORs from sporadic CRC cohorts, there was a difference in the cumulative incidence of CRC for *PMS2* carriers with a PRS2 (based on HRs in this study) in the highest quartile. Bootstrap validation however refuted this promising observation. Further studies are hence needed in other large cohorts.

There were some limitations to this study. Our study consisted exclusively of Dutch *PMS2* carriers and thus had a relatively homogeneous genetic makeup, implying that differences between our results and previous studies might be due to population-specific effects.

A second limitation might be that we did not correct for the specific mutation present in each family, mainly because in the majority of families the segregating *PMS2* mutation is rare or even unique. A previous study by our group did not identify such a correlation with CRC risk in *PMS2* carriers (Supplemental Tables 2a+2b Table: for more details).[34] Unfortunately, we were not able to validate our findings in an external cohort. To our best knowledge this is one of the largest *PMS2* cohorts currently collected, and bootstrap validation is a strong approach to assess discriminative ability of a prediction model.[35] Stratifying our cohort into a discovery and validation cohort was not a viable option as this would have resulted in a substantial decrease in power. Our study might already have been underpowered to detect weak associations. However, while such associations are interesting from a scientific point of view and may be relevant to tumorigenesis, they are not necessarily useful in clinical practice when the effect is small. For the two SNPs previously found to increase risk in *MLH1* mutation carriers, we had 60-80% power to detect an HR of 1.5, which we would consider clinically relevant. The previously reported HR in *MLH1* carriers for rs3802842 was 2.7, an HR for which we have ample power to detect (Supplemental Figure 1).

Families with a segregating *PMS2* mutation show a high degree of phenotypic variability. We were not able to confirm the risk modifying effect of rs3802842 (11q23.1) and s16892766 (8q23.3), which were previously found to increase the risk in *MLH1*-associated LS. This, together with the established lower penetrance, raises the question of whether *PMS2*-associated LS should be considered a separate Lynch disease entity. Additional explanations for phenotypic variability that warrant greater exploration include gene-environment interactions and risk modification by other genetic variants.

REFERENCES

1. Dowty, J.G., A.K. Win, D.D. Buchanan, N.M. Lindor, F.A. Macrae, M. Clendenning, Y.C. Antill, S.N. Thibodeau, G. Casey, S. Gallinger, L.L. Marchand, P.A. Newcomb, R.W. Haile, G.P. Young, P.A. James, G.G. Giles, S.R. Gunawardena, B.A. Leggett, M. Gattas, A. Boussioutas, D.J. Ahnen, J.A. Baron, S. Parry, J. Goldblatt, J.P. Young, J.L. Hopper, and M.A. Jenkins, Cancer risks for MLH1 and MSH2 mutation carriers. Hum.Mutat., 2013. 34(3): p. 490-497.
2. Baglietto, L., N.M. Lindor, J.G. Dowty, D.M. White, A. Wagner, E.B. Gomez Garcia, A.H. Vriends, N.R. Cartwright, R.A. Barnetson, S.M. Farrington, A. Tenesa, H. Hampel, D. Buchanan, S. Arnold, J. Young, M.D. Walsh, J. Jass, F. Macrae, Y. Antill, I.M. Winship, G.G. Giles, J. Goldblatt, S. Parry, G. Suthers, B. Leggett, M. Butz, M. Aronson, J.N. Poynter, J.A. Baron, M.L. Le, R. Haile, S. Gallinger, J.L. Hopper, J. Potter, A. de la Chapelle, H.F. Vasen, M.G. Dunlop, S.N. Thibodeau, and M.A. Jenkins, Risks of Lynch syndrome cancers for MSH6 mutation carriers. J.Natl. Cancer Inst., 2010. 102(3): p. 193-201.
3. ten Broeke, S.W., R.M. Brohet, C.M. Tops, H.M. van der Klift, M.E. Velthuizen, I. Bernstein, G. Capella Munar, E. Gomez Garcia, N. Hoogerbrugge, T.G. Letteboer, F.H. Menko, A. Lindblom, A.R. Mensenkamp, P. Moller, T.A. van Os, N. Rahner, B.J. Redeker, R.H. Sijmons, L. Spruijt, M. Suerink, Y.J. Vos, A. Wagner, F.J. Hes, H.F. Vasen, M. Nielsen, and J.T. Wijnen, Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol, 2015. 33(4): p. 319-25.
4. Ma, X., B. Zhang, and W. Zheng, Genetic variants associated with colorectal cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Gut, 2014. 63(2): p. 326-36.
5. Wijnen, J.T., R.M. Brohet, E.R. van, S. Jagmohan-Changur, A. Middeldorp, C.M. Tops, P.M. van, M.G. Ausems, G.E. Gomez, F.J. Hes, N. Hoogerbrugge, F.H. Menko, T.A. van Os, R.H. Sijmons, S. Verhoef, A. Wagner, F.M. Nagengast, J.H. Kleibeuker, P. Devilee, H. Morreau, D. Goldgar, I.P. Tomlinson, R.S. Houlston, W.T. Van, and H.F. Vasen, Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. Gastroenterology, 2009. 136(1): p. 131-137.
6. Talseth-Palmer, B.A., J.T. Wijnen, I.S. Brenne, S. Jagmohan-Changur, D. Barker, K.A. Ashton, C.M. Tops, T.J. Evans, M. McPhillips, C. Groombridge, J. Suchy, G. Kurzawski, A. Spigelman, P. Moller, H.M. Morreau, W.T. Van, J. Lubinski, H.F. Vasen, and R.J. Scott, Combined analysis of three Lynch syndrome cohorts confirms the modifying effects of 8q23.3 and 11q23.1 in MLH1 mutation carriers. Int.J.Cancer, 2013. 132(7): p. 1556-1564.

7. Win, A.K., J.L. Hopper, D.D. Buchanan, J.P. Young, A. Tenesa, J.G. Dowty, G.G. Giles, J. Goldblatt, I. Winship, A. Boussioutas, G.P. Young, S. Parry, J.A. Baron, D. Duggan, S. Gallinger, P.A. Newcomb, R.W. Haile, M.L. Le, N.M. Lindor, and M.A. Jenkins, Are the common genetic variants associated with colorectal cancer risk for DNA mismatch repair gene mutation carriers? Eur J Cancer, 2013. 49(7): p. 1578-1587.
8. Houlle, S., F. Charbonnier, E. Houivet, J. Tinat, M.P. Buisine, O. Caron, J. Benichou, S. Baert-Desurmont, and T. Frebourg, Evaluation of Lynch syndrome modifier genes in 748 MMR mutation carriers. Eur J Hum Genet, 2011. 19(8): p. 887-92.
9. Senter, L., M. Clendenning, K. Sotamaa, H. Hampel, J. Green, J.D. Potter, A. Lindblom, K. Lagerstedt, S.N. Thibodeau, N.M. Lindor, J. Young, I. Winship, J.G. Dowty, D.M. White, J.L. Hopper, L. Baglietto, M.A. Jenkins, and A. de la Chapelle, The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology, 2008. 135(2): p. 419-428.
10. Umar, A., C.R. Boland, J.P. Terdiman, S. Syngal, A. de la Chapelle, J. Ruschoff, R. Fishel, N.M. Lindor, L.J. Burgart, R. Hamelin, S.R. Hamilton, R.A. Hiatt, J. Jass, A. Lindblom, H.T. Lynch, P. Peltomaki, S.D. Ramsey, M.A. Rodriguez-Bigas, H.F. Vasen, E.T. Hawk, J.C. Barrett, A.N. Freedman, and S. Srivastava, Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J.Natl.Cancer Inst., 2004. 96(4): p. 261-268.
11. Vasen, H.F., I. Tomlinson, and A. Castells, Clinical management of hereditary colorectal cancer syndromes. Nat Rev Gastroenterol Hepatol, 2015. 12(2): p. 88-97.
12. Antoniou, A.C., D.E. Goldgar, N. Andrieu, J. Chang-Claude, R. Brohet, M.A. Rookus, and D.F. Easton, A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. Genet.Epidemiol., 2005. 29(1): p. 1-11.
13. Williams, R.L., A note on robust variance estimation for cluster-correlated data. Biometrics, 2000. 56(2): p. 645-6.
14. Dudbridge, F., Power and predictive accuracy of polygenic risk scores. PLoS Genet, 2013. 9(3): p. e1003348.
15. Clendenning, M., H. Hampel, J. LaJeunesse, A. Lindblom, J. Lockman, M. Nilbert, L. Senter, K. Sotamaa, and A. de la Chapelle, Long-range PCR facilitates the identification of PMS2-specific mutations. Hum.Mutat., 2006. 27(5): p. 490-495.
16. van der Klift, H.M., C.M. Tops, E.C. Bik, M.W. Boogaard, A.M. Borgstein, K.B. Hansson, M.G. Ausems, G.E. Gomez, A. Green, F.J. Hes, L. Izatt, L.P. van Hest, A.M. Alonso, A.H. Vriends, A. Wagner, W.A. van Zelst-Stams, H.F. Vasen, H. Morreau, P. Devilee, and J.T. Wijnen, Quantification of sequence exchange events between

PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients. Hum.Mutat., 2010. 31(5): p. 578-587.

17. van der Klift, H.M., A.R. Mensenkamp, M. Drost, E.C. Bik, Y.J. Vos, H.J. Gille, B.E. Redeker, Y. Tiersma, J.B. Zonneveld, E.G. Garcia, T.G. Letteboer, M.J. Olderode-Berends, L.P. van Hest, T.A. van Os, S. Verhoef, A. Wagner, C.J. van Asperen, S.W. Ten Broeke, F.J. Hes, N. de Wind, M. Nielsen, P. Devilee, M.J. Ligtenberg, J.T. Wijnen, and C.M. Tops, Comprehensive Mutation Analysis of PMS2 in a Large Cohort of Probands Suspected of Lynch Syndrome or Constitutional Mismatch Repair Deficiency (CMMRD) Syndrome. Hum Mutat, 2016.
18. Ingham, S.L., J. Warwick, H. Byers, F. Lalloo, W.G. Newman, and D.G. Evans, Is multiple SNP testing in BRCA2 and BRCA1 female carriers ready for use in clinical practice? Results from a large Genetic Centre in the UK. Clin Genet, 2013. 84(1): p. 37-42.
19. Dunlop, M.G., S.E. Dobbins, S.M. Farrington, A.M. Jones, C. Palles, N. Whiffin, A. Tenesa, S. Spain, P. Broderick, L.Y. Ooi, E. Domingo, C. Smillie, M. Henrion, M. Frampton, L. Martin, G. Grimes, M. Gorman, C. Semple, Y.P. Ma, E. Barclay, J. Prendergast, J.B. Cazier, B. Olver, S. Penegar, S. Lubbe, I. Chander, L.G. Carvajal-Carmona, S. Ballereau, A. Lloyd, J. Vijayakrishnan, L. Zgaga, I. Rudan, E. Theodoratou, J.M. Starr, I. Deary, I. Kirac, D. Kovacevic, L.A. Aaltonen, L. Renkonen-Sinisalo, J.P. Mecklin, K. Matsuda, Y. Nakamura, Y. Okada, S. Gallinger, D.J. Duggan, D. Conti, P. Newcomb, J. Hopper, M.A. Jenkins, F. Schumacher, G. Casey, D. Easton, M. Shah, P. Pharoah, A. Lindblom, T. Liu, C.G. Smith, H. West, J.P. Cheadle, R. Midgley, D.J. Kerr, H. Campbell, I.P. Tomlinson, and R.S. Houlston, Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet, 2012. 44(7): p. 770-6.
20. Abbas, T. and A. Dutta, p21 in cancer: intricate networks and multiple activities. Nat Rev Cancer, 2009. 9(6): p. 400-14.
21. Sinicrope, F.A., G. Roddey, M. Lemoine, S. Ruan, L.C. Stephens, M.L. Frazier, Y. Shen, and W. Zhang, Loss of p21WAF1/Cip1 protein expression accompanies progression of sporadic colorectal neoplasms but not hereditary nonpolyposis colorectal cancers. Clin Cancer Res, 1998. 4(5): p. 1251-61.
22. Loo, L.W.M., M. Lemire, and L. Le Marchand, In silico pathway analysis and tissue specific cis-eQTL for colorectal cancer GWAS risk variants. BMC Genomics, 2017. 18(1): p. 381.
23. Battle, A., C.D. Brown, B.E. Engelhardt, and S.B. Montgomery, Genetic effects on gene expression across human tissues. Nature, 2017. 550(7675): p. 204-213.
24. Barrow, E., J. Hill, and D.G. Evans, Cancer risk in Lynch Syndrome. Fam.Cancer, 2013. 12(2): p. 229-240.

25. Haggar, F.A. and R.P. Boushey, Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg, 2009. 22(4): p. 191-7.
26. Kastrinos, F., H. Uno, C. Ukaegbu, C. Alvero, A. McFarland, M.B. Yurgelun, M.H. Kulke, D. Schrag, J.A. Meyerhardt, C.S. Fuchs, R.J. Mayer, K. Ng, E.W. Steyerberg, and S. Syngal, Development and Validation of the PREMM5 Model for Comprehensive Risk Assessment of Lynch Syndrome. J Clin Oncol, 2017. 35(19): p. 2165-2172.
27. Dashti, S.G., R. Chau, D.A. Ouakrim, D.D. Buchanan, M. Clendenning, J.P. Young, I.M. Winship, J. Arnold, D.J. Ahnen, R.W. Haile, G. Casey, S. Gallinger, S.N. Thibodeau, N.M. Lindor, L. Le Marchand, P.A. Newcomb, J.D. Potter, J.A. Baron, J.L. Hopper, M.A. Jenkins, and A.K. Win, Female Hormonal Factors and the Risk of Endometrial Cancer in Lynch Syndrome. Jama, 2015. 314(1): p. 61-71.
28. Jori, B., R. Kamps, S. Xanthoulea, B. Delvoux, M.J. Blok, K.K. Van de Vijver, B. de Koning, F.T. Oei, C.M. Tops, E.J. Speel, R.F. Kruitwagen, E.B. Gomez-Garcia, and A. Romano, Germ-line variants identified by next generation sequencing in a panel of estrogen and cancer associated genes correlate with poor clinical outcome in Lynch syndrome patients. Oncotarget, 2015. 6(38): p. 41108-22.
29. Movahedi, M., D.T. Bishop, F. Macrae, J.P. Mecklin, G. Moeslein, S. Olschwang, D. Eccles, D.G. Evans, E.R. Maher, L. Bertario, M.L. Bisgaard, M.G. Dunlop, J.W. Ho, S.V. Hodgson, A. Lindblom, J. Lubinski, P.J. Morrison, V. Murday, R.S. Ramesar, L. Side, R.J. Scott, H.J. Thomas, H.F. Vasen, J. Burn, and J.C. Mathers, Obesity, Aspirin, and Risk of Colorectal Cancer in Carriers of Hereditary Colorectal Cancer: A Prospective Investigation in the CAPP2 Study. J.Clin.Oncol., 2015.
30. Win, A.K., J.G. Dowty, D.R. English, P.T. Campbell, J.P. Young, I. Winship, F.A. Macrae, L. Lipton, S. Parry, G.P. Young, D.D. Buchanan, M.E. Martinez, E.T. Jacobs, D.J. Ahnen, R.W. Haile, G. Casey, J.A. Baron, N.M. Lindor, S.N. Thibodeau, P.A. Newcomb, J.D. Potter, L. Le Marchand, S. Gallinger, J.L. Hopper, and M.A. Jenkins, Body mass index in early adulthood and colorectal cancer risk for carriers and non-carriers of germline mutations in DNA mismatch repair genes. Br J Cancer, 2011. 105(1): p. 162-9.
31. Van Duijnhoven, F.J., A. Botma, R. Winkels, F.M. Nagengast, H.F. Vasen, and E. Kampman, Do lifestyle factors influence colorectal cancer risk in Lynch syndrome? Fam.Cancer, 2013. 12(2): p. 285-293.
32. Pande, M., P.M. Lynch, J.L. Hopper, M.A. Jenkins, S. Gallinger, R.W. Haile, L. LeMarchand, N.M. Lindor, P.T. Campbell, P.A. Newcomb, J.D. Potter, J.A. Baron, M.L. Frazier, and C.I. Amos, Smoking and colorectal cancer in Lynch syndrome: results from the Colon Cancer Family Registry and the University of Texas M.D. Anderson Cancer Center. Clin.Cancer Res., 2010. 16(4): p. 1331-1339.

33. Botma, A., F.M. Nagengast, M.G. Braem, J.C. Hendriks, J.H. Kleibeuker, H.F. Vasen, and E. Kampman, Body mass index increases risk of colorectal adenomas in men with Lynch syndrome: the GEOLynch cohort study. J.Clin.Oncol., 2010. 28(28): p. 4346-4353.
34. Suerink, M., H.M. van der Klift, S.W. ten Broeke, O.M. Dekkers, I. Bernstein, G. Capella Munar, E. Gomez Garcia, N. Hoogerbrugge, T.G.W. Letteboer, F.H. Menko, A. Lindblom, A. Mensenkamp, P. Moller, T.A. van Os, N. Rahner, B.J.W. Redeker, M. Olderode, L. Spruijt, Y.J. Vos, A. Wagner, H. Morreau, F.J. Hes, H.F.A. Vasen, C.M. Tops, J.T. Wijnen, and M. Nielsen, The effect of genotypes and parent of origin on cancer risk and age of cancer development in PMS2 mutation carriers. Genet Med, 2016. 18(4): p. 405-409.
35. Steyerberg, E.W., Clinical Prediction Models. Statistics for Biology and Health. 2009: Springer New York.

3

SUPPLEMENTAL METHODS - STATISTICAL ANALYSIS

Follow-up and data-collection

For the vast majority of pre-symptomatically tested family members, i.e. controls, last known age was age at DNA diagnosis and thus – in general - age at start of colonoscopic surveillance. This is due to the fact that carriers were ascertained through genetic centres, which made us unable to collect follow-up data, as subsequent surveillance was done at the gastroenterology and/or gynaecology departments.

Ascertainment bias & weighted analysis

All subjects included in this study were derived from family cancer clinics and were therefore not randomly selected with respect to their phenotype. In other words they were selected based on the occurrence of cancer at a young age or due to several family members being affected. Therefore, these carriers usually belong to high-risk families, ascertained as a consequence of their relatively severe phenotype. It is likely that other factors than the germline mutation in PMS2 and the SNPs currently investigated may play a role in the phenotype variability, such as lifestyle or other genetic factors.[1]

Weights were calculated based on incidence rates in the Dutch population. HRs based on a proportion of this cohort have been previously reported and were used to determine age stratum (5 year) specific weights.[3] All calculated weights for cases were smaller than 1, effectively down-weighting cases compared with controls. It is important to note that for hypothesis testing the unweighted p-value and confidence interval are reliable.[12] We therefore report the weighted HRs with both the p-values and the 95% confidence intervals from the unweighted analysis. Both unweighted and weighted HRs are listed in the tables.

Polygenic risk score

PRS1 and 2 included all 24 SNPs that were found to be in Hardy Weinberg equilibrium, however two SNPs in the HFE gene were not taken into account in the PRS1 calculation, because to our knowledge there were no ORs from meta-analysis reported in current literature. The PRS was calculated as previously described[16], by using the following formula for PRS1 and PRS 2 respectively:

$$\sum_{i=1}^{n} a_i \log OR_i$$

$$\sum_{i=1}^{n} a_i \log HR_i$$

where n is the number of SNPs, a is the number of risk alleles for each SNP and the OR the meta-analysis derived OR (supplementary table 1)[4] or the HR from the current study for each SNP for PRS1 and PRS 2 respectively. *PMS2* carriers were categorized into four groups of equal sizes based on the quartiles in the control group.

Multiple testing

Correction for multiple testing was done by calculating the corrected overall critical p-value for all performed tests (four per SNP, including gender stratification, four PRS tests and the combination of rs3802842 and rs16892766) using the Bonferroni method. This leads to a p value 0.05/101= 0.0005.

Other

Mean age at CRC development was examined using one-way Analysis of Variance (ANOVA).

SUPPLEMENTARY TABLE 1 Overview of all tested SNPs, including gender stratification

SNP / OR from literature	Alleles (minor allele in bold)	MAF (dbSNP Overall/ european)	Allele frequencies *PMS2* cohort					Total genotyped for SNP (total n=507)
			Cases (CRC)		Controls			
			Genotype frequency	MAF	Genotype frequency	MAF	p	
rs6687758	AA	1000 genomes: G=0.19	76	0.21	262	0.17	0.36	502
1q41	A**G**	European: 0.22	43		108			
DUSP10	**GG**		3		10			
OR: 1.09 (Ma)	AG+GG							
	Per allele							
rs6691170	GG	1000 genomes: T=0.26	40	0.41	143	0.38	0.61	500
1q41	G**T**	European: 0.40	64		182			
DUSP10	**TT**		17		54			
OR: 1.06 (Ma)	Per allele							
rs10936599	**TT**	ExAC: T=0.28	7	0.25	24	0.26	0.96	495
3q26.6	T**C**	1000 genomes: T=0.27	48		146			
MYNN	CC	European: 0.24	67		203			
OR: 0.93	Per allele							
rs1321311	CC	1000 genomes A=0.28	65	0.26	234	0.23	0.19	503
6p21.2	C**A**	European: 0.22	50		122			
CDKN1A	**AA**		7		25			
OR: 1.1	Per allele							
rs16892766	AA	1000 genomes: C=0.08	88	0.14	290	0.13	0.67	502
8q23.3	**CA**	European: 0.09	30		85			
EIF3H	**CC**		3		6			
OR: 1.25 (Ma)	CA+CC							

Cox regression - overall			Cox regression - Males				Cox regression - females			
HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p
		0.6								
1.20 (0.82-1.76)	1.27	0.4								
0.98 (0.32-2.97)	0.97	1								
1.18 (0.82-1.72)	1.24	0.4								
1.13 (0.82-1.55)	1.17	0.5	192 (59)	1.21 (0.72-2.03)	1.47	0.5	310 (63)	1.08 (0.71-1.65)	1.07	0.7
		0.7								
1.16 (0.76-1.75)	1.18	0,5								
0.95 (0.51-1.76)	0.97	0.9								
1.01 (0.77-1.33)	1.03	0.9	191 (58)	0.97 (0.66-1.44)	1.05	0.9	309 (63)	1.06 (0.74-1.53)	0.9	0.7
		0.8								
0.73 (0.29-1.88)	0.77	0.5								
0.72 (0.27-1.91)	0.76	0.5								
0.92 (0.65-1.32)	0.93	0.7	189 (60)	0.79 (0.44-1.41)	0.7	0.4	306 (62)	1.03 (0.69-1.54)	1.38	0.9
		0.1								
1.45 (1.00-2.12)	1.57	0.1								
1.02 (0.50-2.05)	1.13	1								
1.19 (0.92-1.56)	1.28	0.2	193 (60)	1.90 (1.21-2.96)	2.07	0	310 (62)	0.90 (0.63-1.28)	0.83	0.6
		0.9								
1.12 (0.73-1.72)	0.99	0.6								
1.11 (0.34-3.57)	1.6	0.9								
1.12 (0.73-1.71)	1.04	0.6								

SUPPLEMENTARY TABLE 1 Overview of all tested SNPs, including gender stratification

SNP / OR from literature	Alleles (minor allele in bold)	MAF (dbSNP Overall/ european)	Allele frequencies *PMS2* cohort					Total genotyped for SNP (total n=507)
			Cases (CRC)		Controls			
			Genotype frequency	MAF	Genotype frequency	MAF	p	
	Per allele							
rs6983267	**TT**	1000 genomes: T=0.39	31	0.48	108	0.53	0.27	485
8q24.21	**TG**	European: 0.50	49		174			
MYC	GG		36		87			
OR: 1.21 (Ma)	Per allele							
rs10795668	**AA**	1000 genomes: A=0.23	12	0.29	31	0.29	0.82	502
10p14	**AG**	European: 0.32	49		160			
FLJ3802842	GG		61		189			
OR: 0.89 (Ma)	Per allele							
rs3802842	AA	1000 genomes: C=0.29	67	0.25	190	0.28	0.79	495
11q23	**AC**	European: 0.27	49		159			
COLCA1/ COLCA2	**CC**		6		24			
OR: 1.11 (Ma)	Per allele							
rs3824999	**AA**	1000 genomes: G=0.33	24	0.47	89	0.49	0.55	498
11q13.4	**AC**	European: 0.52	67		187			
POLD3	CC		31		100			
OR: 0.93 (Ma)	Per allele							
rs4444235	TT	1000 genomes: C=0.43	44	0.39	105	0.47	0.1	497
14q22.2	**TC**	European: 0.49	59		188			

Cox regression - overall			Cox regression - Males				Cox regression - females			
HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p
1.09 (0.76-1.59)	1.08	0.6	191 (58)	0.91 (0.57-1.45)	0.82	0.7	311 (63)	1.19 (0.70-2.01)	0.9	0.5
		0.8								
1.04 (0.61-1.78)	1.1	0.9								
1.16 (0.67-2.01)	1.19	0.6								
1.07 (0.82-1.42)	1.09	0.6	185 (57)	1.09 (0.74-1.60)	1.03	0.7	300 (59)	1.10 (0.79-1.54)	1.24	0.6
		0.4								
0.74 (0.34-1.59)	0.66	0.4								
0.95 (0.46-1.97)	0.81	0.9								
1.09 (0.81-1.47)	1.02	0.6	191 (59)	0.97 (0.66-1.43)	0.77	0.9	311 (63)	1.25 (0.78-2.00)	1.15	0.4
		0.7								
0.85 (0.60-1.22)	0.81	0.4								
1.02 (0.47-2.22)	0.81	1								
0.92 (0.67-1.25)	0.84	0.6	189 (60)	0.73 (0.44-1.22)	0.74	0.2	307 (63)	1.08 (0.68-1.70)	1.18	0.8
		0.3								
1.41 (0.93-2.15)	1.28	0.1								
1.20 (0.71-2.03)	1.12	0.5								
1.07 (0.85-1.37)	1.05	0.5	192 (60)	1.10 (0.75-1.62)	1.28	0.6	306 (62)	1.04 (0.77-1.39)	1	0.8
		0.2								
0.82 (0.55-1.25)	0.89	0.4								

SUPPLEMENTARY TABLE 1 Overview of all tested SNPs, including gender stratification

SNP / OR from literature	Alleles (minor allele in bold)	MAF (dbSNP Overall/ european)	Allele frequencies *PMS2* cohort					Total genotyped for SNP (total n=507)
			Cases (CRC)		Controls			
			Genotype frequency	MAF	Genotype frequency	MAF	p	
BMP4	**CC**		18		83			
OR: 1.11 (Ma)	Per allele							
rs9929218	**AA**	1000 genomes: A=0.26	8	0.30	31	0.29	0.54	502
16q22.1	**AG**	European: 0.29	57		155			
CDH1	GG		58		193			
OR: 0.91 (Ma)	Per allele							
rs4939827	**CC**	1000 genomes: T=0.35	22	0.48	82	0.49	0.51	500
18q21.1	**CT**	European: 0.53	73		204			
SMAD7	TT		27		92			
OR: 0.85 (Ma)	Per allele							
rs12953717	CC	1000 genomes: T=0.30	36	0.42	120	0.44	0.83	500
18q21.1	**CT**	European: 0.45	65		196			
SMAD7	**TT**		22		61			
OR: 1.12	Per allele							
rs10411210	**TT**	1000 genomes: T=0.26	1	0.09	3	0.10	0.96	489
19q13.1	**TC**	European: 0.10	20		65			
RHPN2	CC		100		300			
OR: 0.87 (Ma)	CC (ref)							
	CT+TT							
rs961253	CC	1000 genomes: A=0.29	43	0.37	149	0.37	0.49	495
		European:						

Cox regression - overall			Cox regression - Males				Cox regression - females			
HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p
0.60 (0.35-1.03)	0.54	0.1								
0.78 (0.61-1.01)	0.77	0.1	192 (59)	0.80 (0.51-1.26)	0.76	0.3	305 (62)	0.74 (0.53-1.05)	0.85	0.1
		0.7								
1.07 (0.51-2.30)	1.51	0.9								
0.93 (0.41-2.11)	1.26	0.9								
0.91 (0.66-1.26)	0.96	0.6	193 (60)	0.92 (0.57-1.47)	0.97	0.7	309 (63)	0.97 (0.63-1.50)	0.85	0.9
		0.7								
1.19 (0.73-1.96)	1.31	0.5								
1.00 (0.57-1.78)	1.08	1								
0.99 (0.77-1.29)	1.02	1	192 (59)	0.91 (0.61-1.36)	0.74	0.6	308 (63)	1.02 (0.73-1.44)	1.04	0.9
		0.8								
1.08 (0.73-1.63)	1.15	0.7								
1.17 (0.73-1.88)	1.29	0.5								
1.08 (0.86-1.37)	1.14	0.5	192 (60)	1.07 (0.75-1.54)	1.09	0.7	308 (63)	1.04 (0.74-1.47)	1.03	0.8
		0.2								
0.26 (0.064-1.09)		0.1								
0.32 (0.087-1.18)		0.1								
0.86 (0.51-1.44)	0.82	0.6	190 (60)	1.34 (0.61-2.94)	1.36	0.5	299 (61)	1.00 (0.52-1.94)	1.07	1
		0.9								

SUPPLEMENTARY TABLE 1 Overview of all tested SNPs, including gender stratification

SNP / OR from literature	Alleles (minor allele in bold)	MAF (dbSNP Overall/ european)	Allele frequencies *PMS2* cohort					Total genotyped for SNP (total n=507)
			Cases (CRC)		Controls			
			Genotype frequency	MAF	Genotype frequency	MAF	p	
20p12.3	C**A**	0.36	63		179			
BMP2	**AA**		12		49			
OR: 1.12 (Ma)	Per allele							
rs1569686	GG	1000 genomes: G=0.28	49	0.39	144	0.38	0.38	499
20q11.2	G**T**	European: G=0.63	52		181			
DNMT3B	**TT**		22		51			
OR: 0.57 (Ma)	Per allele							
rs2736100	**GG**	1000 genomes: C=0.48	31	0.50	90	0.46	0.36	498
5p15	G**T**	European: 0.50	60		168			
TERT	TT		30		119			
OR: 1.07 (Ma)	Per allele							
rs1800734	GG	1000 genomes: A=0.32	77	0.22	245	0.20	0.88	501
3p21.3	G**A**	European: 0.27	38		112			
MLH1	**AA**		8		21			
OR: 1.51 (Ma)	Per allele							
rs1799945	CC	ExAC: G=0.11	93	0.13	285	0.13	0.74	502
6p21.3	C**G**	1000 genomes: G=0.07	26		89			
HFE	**GG**	European: 0.17	3		6			
No meta-analysis available	CG+GG							
	Per allele							
Rs5934683	CC	1000 genomes: C=0.34						

Cox regression - overall			Cox regression - Males				Cox regression - females			
HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p
1.08 (0.73-1.58)	1.09	0.7								
1.06 (0.62-1.83)	1.27	0.8								
1.04 (0.81-1.34)	1.12	0.7	191 (58)	1.20 (0.82-1.75)	1.25	0.3	304 (60)	0.87 (0.60-1.26)	0.88	0.5
		0.5								
0.96 (0.65-1.42)		0.8								
1.26 (0.77-2.06)		0.4								
1.09 (0.85-1.40)	1.06	0.5	191 (60)	1.11 (0.79-1.55)	1.08	0.6	308 (63)	1.04 (0.72-1.49)	1.23	0.8
		0.1								
1.09 (0.68-1.75)	1.26	0.7								
0.73 (0.42-1.26)	0.8	0.3								
0.86 (0.66-1.10)	0.89	0.2	192 (60)	0.69 (0.48-0.97)	0.6	0	306 (61)	0.99 (0.71-1.38)	1.03	1
		0.7								
0.99 (0.69-1.43)	1.14	1								
1.30 (0.67-2.54)	1.62	0.4								
1.06 (0.81-1.41)	1.21	0.7	193 (60)	0.80 (0.51-1.27)	0.81	0.3	308 (63)	1.25 (0.86-1.81)	1.89	0.2
		0.4								
1.27 (0.84-1.93)	1.35	0.3								
1.76 (0.45-6.93)	1.36	0.4								
1.31 (0.86-1.99)	1.35	0.2								
1.29 (0.87-1.92)	1.29	0.2	192 (59)	1.58 (0.92-2.70)	2.07	0.1	310 (63)	1.12 (0.64-1.95)	1.41	0.7
										0.2

SUPPLEMENTARY TABLE 1 Overview of all tested SNPs, including gender stratification

SNP / OR from literature	Alleles (minor allele in bold)	MAF (dbSNP Overall/ european)	Allele frequencies *PMS2* cohort					Total genotyped for SNP (total n=507)
			Cases (CRC)		Controls			
			Genotype frequency	MAF	Genotype frequency	MAF	P	
Xp22.2	**TC**	European: 0.46						
SHROOM2	**TT**	GoNL: C=0.63						
OR: 1.07 (Ma)	Per allele							
Rs1800562	CC	ExAC=0.03	106	0.07	334	0.06	0.82	498
6p21.3	**TC**	1000 genomes: A=0.01	15		40			
HFE	**TT**	European: 0.04	1		2			
No meta-analysis available	TC+TT							
	Per allele							
rs11169552	**TT**	1000 genomes: T=0.25	5	0.25	25	0.26	0.55	499
12q13.3	**TC**	European: 0.25	51		146			
ATF1	CC		66		206			
OR: 0.92 (Ma)	Per allele							
12q13.13	CC	1000 genomes: T=0.50	47	0.39	162	0.34	0.34	501
LARP4/ DIP2B	**TC**	European: 0.35	56		176			
OR: 1.06 (Ma)	**TT**		19		41			
	Per allele							
rs4779584	CC	1000 genomes: T=0.49	75	0.19	246	0.22	0.69	499
15q13.3	**TC**	European: 0.20	41		117			
GREM1/ SCG5	**TT**		6		14			
OR: 1.26 (Ma)	Per allele							

Threshold for multiple testing: p<0.0005. HR=Hazard ratio. OR=Odds ratio. 95%CI= 95% confidence interval.

Cox regression - overall			Cox regression - Males				Cox regression - females			
HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p	PMS2 carriers (cases)	HR (95% CI)	wHR	p
N/A								1.43 (0.87-2.35)	0.69	0.2
								0.62 (1.19-2.03)	0.73	0.4
			191 (59)	1.08 (0.78-1.50)	1.07	0.6	310 (63)	1.05 (0.74-1.51)	0.78	0.8
		1								
1.05 (0.56-1.96)	1.18	0.9								
1.03 (0.093-11.4)	0.8	1								
1.05 (0.57-1.93)	1.14	0.9								
1.04 (0.60-1.83)	1.1	0.9	190 (59)	0.99 (0.49-1.97)	1.36	1	308 (63)	1.04 (0.49-2.21)	0.6	0.9
		0.5								
1.60 (0.69-3.71)	2.4	0.3								
1.48 (0.62-3.57)	1.96	0.4								
1.04 (0.77-1.4)	1.03	0.8	190 (59)	1.20 (0.80-1.82)	1.13	0.4	309 (63)	0.88 (0.58-1.33)	0.75	0.5
		0.6								
1.11 (0.75-1.66)	1.19	0.6								
1.36 (0.78-2.36)	1.46	0.3								
1.15 (0.88-1.51)	1.2	0.3	190 (59)	1.44 (0.99-2.10)	1.4	0.1	311 (63)	0.94 (0.67-1.32)	0.94	0.7
		0.4								
1.12 (0.76-1.64)	1.07	0.6								
1.88 (0.66-5.39)	2.5	0.2								
1.21 (0.88-1.67)	1.28	0.2	191 (60)	1.82 (1.16-2.86)	1.81	0	308 (62)	0.89 (0.57-1.39)	0.83	0.6

MAF=minor allele frequency. wHR: weighted Hazard Ratio. #This meta-analysis uses homozygotes for the major allele as a reference category.

SUPPLEMENTARY TABLE 2A Overview of germline *PMS2* mutations in the cohort. Each row describes 1 family

Genotype molecular	Genotype aminoacid	no crc	crc	Total
c.856_857delGA	p.Asp286Glnfs*12	0	1	1
c.1078_1081dupATAG	p.Gly361Aspfs*5	4	0	4
c.1079_1080del	p.Ile360Argfs*4	3	1	4
c.1144+2T>A	p.(Glu330_Glu381del)	3	0	3
c.1145-?_(*160_?)del	deletion exon 11-15	11	1	12
c.1214C>A	p.Ser405*	3	1	4
c.1261C>T	p.Arg421*	1	1	2
c.137G>T	p.Ser46Ile	5	2	7
c.137G>T	p.Ser46Ile	5	1	6
c.137G>T	p.Ser46Ile	5	0	5
c.137G>T	p.Ser46Ile	3	1	4
c.137G>T	p.Ser46Ile	2	0	2
c.137G>T	p.Ser46Ile	2	0	2
c.137G>T	p.Ser46Ile	0	1	1
c.137G>T	p.Ser46Ile	0	1	1
c.137G>T	p.Ser46Ile	0	1	1
c.137G>T	p.Ser46Ile	0	1	1
c.163+2T>C	p.SerArgfs*5	2	1	3
c.164-?_803+?del	deletion exon 3-7	1	0	1
c.1730dup	p.Arg578Alafs*	1	0	1
c.1831dupA	p.Ile611Asnfs*2	3	1	4
c.1831dupA	p.Ile611Asnfs*2	1	0	1
c.1831dupA	p.Ile611Asnfs*2	0	1	1
c.1831dupA	p.Ile611Asnfs*2	2	1	3
c.1882C>T	p.Arg628*	13	2	15
c.1882C>T	p.Arg628*	7	2	9
c.1882C>T	p.Arg628*	8	1	9
c.1882C>T	p.Arg628*	6	1	7
c.1882C>T	p.Arg628*	4	2	6
c.1882C>T	p.Arg628*	5	0	5
c.1882C>T	p.Arg628*	3	2	5

SUPPLEMENTARY TABLE 2A Overview of germline *PMS2* mutations in the cohort. Each row describes 1 family

Genotype molecular	Genotype aminoacid	no crc	crc	Total
c.1882C>T	p.Arg628*	2	1	3
c.1882C>T	p.Arg628*	2	1	3
c.1882C>T	p.Arg628*	3	0	3
c.1882C>T	p.Arg628*	2	1	3
c.1882C>T	p.Arg628*	2	1	3
c.1882C>T	p.Arg628*	1	1	2
c.1882C>T	p.Arg628*	1	1	2
c.1882C>T	p.Arg628*	1	0	1
c.1882C>T	p.Arg628*	0	1	1
c.1882C>T	p.Arg628*	0	1	1
c.1882C>T	p.Arg628*	0	1	1
c.211_214del	p.Asn71Aspfs*4	1	1	2
c.211_214del	p.Asn71Aspfs*4	0	1	1
c.2117delA	p.Lys706Serfs*19	3	1	4
c.2117delA	p.Lys706Serfs*19	2	0	2
c.2155C>T	p.Gln719*	1	1	2
c.2155C>T	p.Gln719*	0	1	1
c.2174+1G>A	p.?	2	0	2
c.219_220dup	p.Gly74Valfs*3	5	2	7
c.219_220dup	p.Gly74Valfs*3	5	2	7
c.219_220dup	p.Gly74Valfs*3	6	0	6
c.219_220dup	p.Gly74Valfs*3	5	0	5
c.219_220dup	p.Gly74Valfs*3	1	0	1
c.2192_2196delTAACT	p.Leu731Cysfs*3	4	2	6
c.2192_2196delTAACT	p.Leu731Cysfs*3	5	0	5
c.2192_2196delTAACT	p.Leu731Cysfs*3	4	1	5
c.2192_2196delTAACT	p.Leu731Cysfs*3	4	0	4
c.2192_2196delTAACT	p.Leu731Cysfs*3	2	1	3
c.2192_2196delTAACT	p.Leu731Cysfs*3	2	1	3
c.2192_2196delTAACT	p.Leu731Cysfs*3	0	2	2
c.2192_2196delTAACT	p.Leu731Cysfs*3	0	1	1

SUPPLEMENTARY TABLE 2A Overview of germline *PMS2* mutations in the cohort. Each row describes 1 family

Genotype molecular	Genotype aminoacid	no crc	crc	Total
c.2275+1G>A	p.?	4	0	4
c.2275+1G>A	p.?	3	1	4
c.2404C>T	p.Arg802*	0	1	1
c.24-12_107del96bp	intronic	2	0	2
c.24-12_107delinsAAAT	p.?/p.Ser8Argfs*5	4	1	5
c.2444C>T	p.Ser815Leu	4	1	5
c.2445+1G>T	p.?	0	1	1
c.247_250dup	p.Thr84Ilefs*9	2	0	2
c.251-2A>C	p.?	2	0	2
c.251-2A>C	p.?	1	1	2
c.251-2A>C	p.?	0	1	1
c.251-2A>C	p.?	0	1	1
c.2T>A		1	0	1
c.319C>T	p.Arg107Trp	6	1	7
c.325dupG	p.Glu109Glyfs*30	3	1	4
c.325dupG	p.Glu109Glyfs*30	3	0	3
c.325dupG	p.Glu109Glyfs*30	0	1	1
c.325dupG	p.Glu109Glyfs*30	0	1	1
c.658dupG	p.Ser220fs	0	1	1
c.697C>T	p.Gln233*	5	0	5
c.697C>T	p.Gln233*	1	1	2
c.697C>T	p.Gln233*	2	0	2
c.697C>T	p.Gln233*	1	0	1
c.697C>T	p.Gln233*	0	1	1
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	7	1	8
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	3	2	5
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	3	1	4
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	3	1	4
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	3	1	4
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	4	0	4
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	2	2	4

SUPPLEMENTARY TABLE 2A Overview of germline *PMS2* mutations in the cohort. Each row describes 1 family

Genotype molecular	Genotype aminoacid	no crc	crc	Total
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	4	0	4
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	3	0	3
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	2	1	3
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	1	1	2
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	0	2	2
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	1	1	2
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	0	1	1
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	0	1	1
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	1	0	1
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	0	1	1
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	0	1	1
c.736_741delCCCCCTinsTGTGTGTGAAG	p.Pro246Cysfs*3	1	0	1
c.804-?_903+?del	p.Tyr268*	0	1	1
c.804-60_804-59ins2kb	p.?	3	1	4
c.804-60_804-59ins2kb	p.?	0	1	1
c.804-60_804-59ins2kb	p.?	1	0	1
c.804-60_804-59ins2kb	p.?	0	1	1
c.823C>T	p.Gln275*	5	1	6
c.861_864delACAG	p.Arg287Serfs*19	4	0	4
c.903G>T	p.Tyr268*	0	1	1
c.904-?_988+?dup	p.Glu330Glyfs*7	4	0	4
c.904-?_988+?dup	p.Glu330Glyfs*7	4	0	4
c.904_911delGTCTGCAG	p.Val302Thrfs*4	1	0	1
c.943C>T	p.Arg315*	2	0	2
c.989-?_2275+?del		6	2	8
c.989_1144 + 685del		1	0	1
deletion at least exon 1-11		0	1	1
deletion at least exon 9-11		2	0	2
deletion entire PMS2		3	1	4
deletion entire PMS2		2	0	2
deletion exon 10		4	0	4

SUPPLEMENTARY TABLE 2A Overview of germline *PMS2* mutations in the cohort. Each row describes 1 family

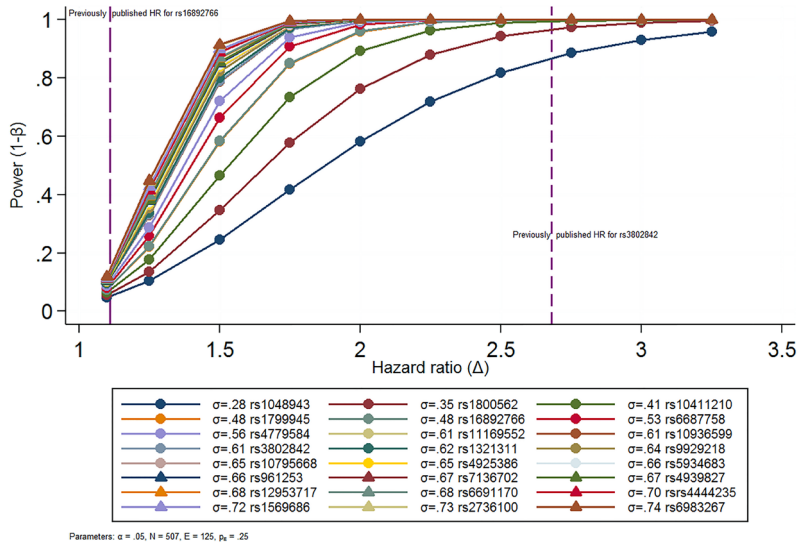
Genotype molecular	Genotype aminoacid	no crc	crc	Total
deletion exon 10		1	1	2
deletion exon 10		0	1	1
deletion exon 11-15		6	1	7
deletion exon 11-15		5	2	7
deletion exon 11-15		4	1	5
deletion exon 11-12		6	0	6
deletion exon 11-12		4	0	4
deletion exon 11-12		2	0	2
deletion exon 11-12		1	1	2
deletion exon 13-15		2	0	2
c.2276-?_c.2445+?del	deletion exon 14	5	0	5
c.2276-?_c.2445+?del		2	2	4
c.2276-?_c.2445+?del		1	1	2
c.2276-?_c.2445+?del		0	1	1
deletion exon 14 and 15		1	0	1
deletion exon 2		10	4	14
deletion exon 2		8	1	9
deletion exon 2		2	1	3
deletion exon 2, 5-11, 13-15		0	1	1
deletion exon 2-4 (c.24-?_353+?del)		3	2	5
deletion exon 5-15		0	1	1
deletion exon 5-7		5	1	6
deletion exon 5-7		4	2	6
deletion exon 5-7		1	1	2
deletion exon 5-7		0	1	1
deletion exon 6-7		0	1	1
duplication of exon 11 + exon 12, inserted in intron 12		0	1	1
Two segregating mutations: deletion PMS2 exon 1 t/m 11 OR deletion exon 5 t/m 7		17	1	18
Total		**383**	**124**	**507**

SUPPLEMENTARY TABLE 2B Most frequent germline *PMS2* mutations in the cohort

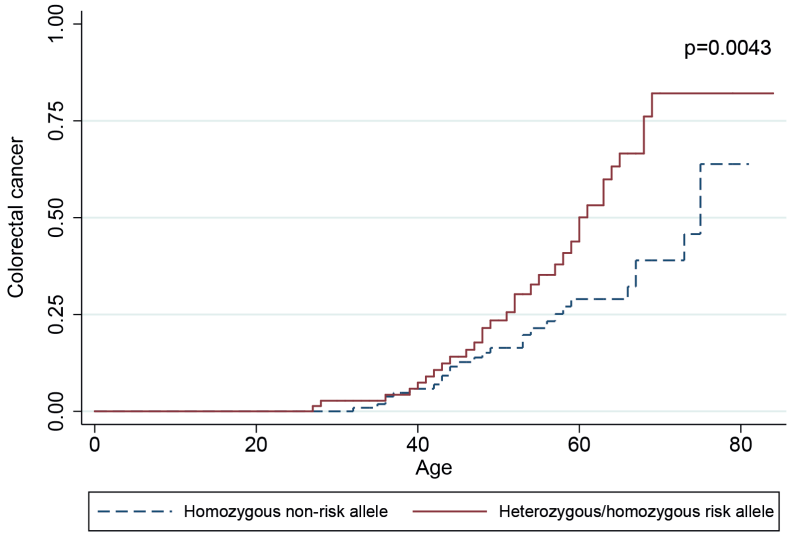
GENOTYPE	No CRC			CRC	Total	Mean members per family	Mean affected members per family	Mean age at CRC for cases	Classification Suerink et al
c.1145-?_(*160_?)del	11	1	12	2 (17%)	1	12,0	1,0	45,0	3
c.137G>T	19	8	27	12 (44%)	10	2,7	0,8	58,7	2
c.1882C>T	60	19	79	24 (30%)	18	4,4	1,1	55,2	1
c.2192_2196delTAACT	21	8	16	12 (41%)	8	2,0	1,0	53,5	1
c.219_220dup	22	4	26	8 (31%)	5	5,2	0,8	65,0	1
c.697C>T	9	2	11	5 (45%)	5	2,2	0,4	51,5	1
c.736_741delCCCCCTinsTGTGTGTGAAG	40	17	57	20 (35%)	20	2,9	0,9	50,3	1
deletion exon 11-15	15	4	19	9 (47%)	3	6,3	1,3	58,8	1
deletion exon 11-12	13	1	14	9 (64%)	5	2,8	0,2	65,0	3
deletion exon 14	8	3	11	6 (55%)	3	3,7	1,0	53,8	3
deletion exon 2	20	6	26	10 (38%)	3	8,7	2,0	57,2	3
deletion exon 5-7	19	6	25	13 (52%)	5	5,0	1,2	55,5	3
Total	**252**	**77**	**316**	**129 (41%)**	86				

CRC: colorectal cancer
A previous study by our group (Suerink et al[1]) showed that mutation carriers with loss of RNA expression (group 1) had lower age at CRC diagnosis (difference 10 years), compared to group 2 mutation carriers. We do not observe marked differences in mean age at CRC diagnosis for this group. Nor did we see some mutations with a high affected family member amount. This together with the results of the study by Suerink et al suggests that genotype did not heavily influence our results.
1=no mRNA expression from mutated allele, 2=normal mRNA expression; 3=RNA expression unknown

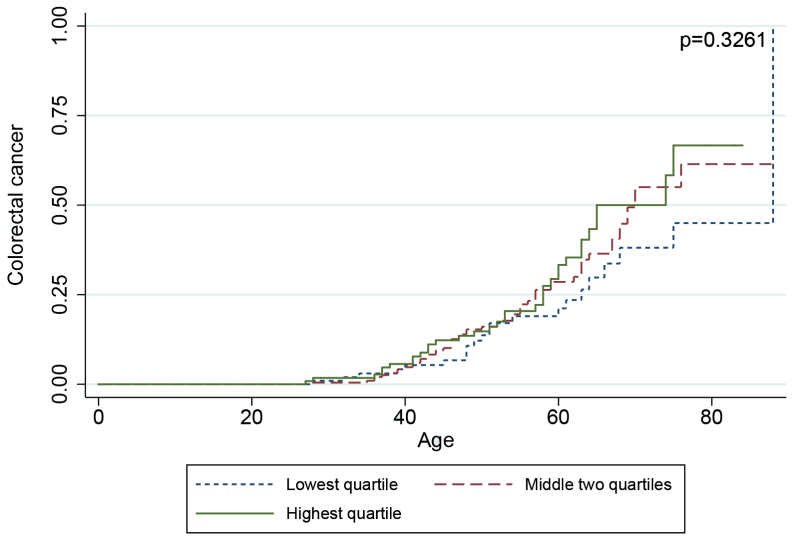
1. Suerink M, van der Klift HM, ten Broeke SW, et al. The effect of genotypes and parent of origin on cancer risk and age of cancer development in PMS2 mutation carriers. Genet Med 2016;18:405-409.



SUPPLEMENTARY FIGURE 1 Post-hoc power analysis.



SUPLEMENTARY FIGURE 2 Kaplan Meier survival curves with endpoint colorectal cancer for Rs1321311. The p-value represents the comparison between PMS2 carriers who are homozygous for the non-risk allele with PMS2 carriers who are heterozygous/homozygous for the risk allele.



SUPPLEMENTARY FIGURE 3 Kaplan Meier survival curves with endpoint colorectal cancer for polygenic risk score (PRS) 2. The p-value represents the comparison for the lowest, the two middle and the highest quartile of the PRS.