

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2010 June ; 19(6): 1478–1483. doi:10.1158/1055-9965.EPI-09-1320.

Low-Penetrance Susceptibility Variants in Familial Colorectal Cancer

Iina Niittymäki¹, Eevi Kaasinen¹, Sari Tuupanen¹, Auli Karhu¹, Heikki Järvinen², Jukka-Pekka Mecklin³, Ian P.M. Tomlinson⁴, Maria Chiara Di Bernardo⁵, Richard S. Houlston⁵, and Lauri A. Aaltonen¹

¹ Department of Medical Genetics, Genome-Scale Biology Research Program, Biomedicum Helsinki, University of Helsinki, Finland ² Department of Surgery, Helsinki University Hospital, Helsinki, Finland ³ Department of Surgery, Jyväskylä Central Hospital, Jyväskylä, Finland ⁴ Molecular and Population Genetics, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom ⁵ Section of Cancer Genetics, Institute of Cancer Research, Sutton, United Kingdom

Abstract

Background: Genome-wide association studies have identified ten low-penetrance loci that confer modestly increased risk of colorectal cancer (CRC). Although they underlie a significant proportion of CRC in the general population, their impact on the familial risk of CRC has yet to be formally enumerated. The aim of this study was to examine the combined contribution of the ten variants, rs6983267, rs4779584, rs4939827, rs16892766, rs10795668, rs3802842, rs4444235, rs9929218, rs10411210, and rs961253, on familial CRC.

Methods: The population-based series of CRC samples included in this study consisted of 97 familial cases and 691 sporadic cases. Genotypes in the ten loci and clinical data, including family history of cancer verified from the Finnish Cancer Registry, were available. The overall number of risk alleles (0–20) was determined and its association with familial CRC analyzed. Excess familial risk was estimated using cancer incidence data from the first-degree relatives of the cases.

Results: A linear association between the number of risk alleles and familial CRC was observed ($P=0.006$). With each risk allele addition, the odds of having an affected first-degree relative increased by 1.16 (95% confidence interval 1.04–1.30). The ten low-penetrance loci collectively explain approximately 9% of the variance in familial risk of CRC.

Conclusions: This study provides evidence to support the previous indirect estimations that these low-penetrance variants account for a relatively small proportion of the familial aggregation of CRC.

Impact: Our results emphasize the need to characterize the remaining molecular basis of familial CRC, which should eventually yield in individualized targeting of preventive interventions.

Keywords

familial colorectal cancer; genome-wide association study; single nucleotide polymorphism; low-penetrance; association

Correspondence and requests for reprints: Lauri A. Aaltonen, Address: Department of Medical Genetics, Biomedicum Helsinki, PO Box 63, FIN-00014 University of Helsinki, Finland, Fax: +358-9-19125105, lauri.aaltonen@helsinki.fi.

Websites

Primer3 software: <http://frodo.wi.mit.edu/primer3/>

Ensembl database: www.ensembl.org

Introduction

Family history is a major risk factor for colorectal cancer (CRC) with individuals with an affected first-degree relative having a two-fold increased risk (1). Concordance between mono- and dizygotic twins indicate hereditary factors to underlie approximately 35% of CRC (2). However, germline mutations in high-penetrance genes, including *MLH1*, *MSH2*, *APC*, and *MYH* account for only ~5% of all CRCs (3). According to the “common disease - common variant” concept, much of the remaining inheritance is likely caused by a large number of low-penetrance variants that are common in the population. Genome-wide association (GWA) studies conducted in UK and Canada, in which large case-control sets have been genotyped for single nucleotide polymorphisms (SNPs), have so far identified ten chromosomal loci that confer a modest risk of CRC. The first CRC-associating variant to emerge in 2007 was rs6983267 at 8q24 that increases the risk of CRC with an odds ratio (OR) of 1.21 (95% CI 1.15-1.27) (4-6). Subsequently, five more loci at 18q21, 15q13, 8q23, 10p14, and 11q23 were published in which the most strongly associating SNPs were rs4939827, rs4779584, rs16892766, rs10795668, and rs3802842, respectively (7-11). In the first phases of the GWA studies, less than thousand cases and controls were genotyped with microarrays containing half a million SNPs, but all the six associations were later replicated in around 7,000-15,000 cases and controls. After the first round, additional four variants, rs4444235 at 14q22, rs9929218 at 16q22, rs10411210 at 19q13, and rs961253 at 20p12 were identified through meta-analysis where these associations were replicated in more than 20,000 cases and controls (12).

All the ten low-penetrance loci independently predispose to CRC with allelic odds ratios of 1.10-1.26 (95% CI range 1.06-1.34) and have risk allele frequencies of 0.07-0.90 in the general population (13). Although typically the associating tagging SNPs are merely correlated with the actual causal variants, we have shown that rs6983267 at 8q24 affects Wnt signalling by disrupting an enhancer element and possibly driving the expression of the 335 kb distal *MYC* oncogene (14). In 18q21 the functional change is a novel variant altering *SMAD7* expression (15). Indeed, all the ten tagging SNPs locate in either intergenic or intronic areas of the genome and many might therefore affect gene expression through distant regulatory elements. Aberrant transforming growth factor beta (TGF β) superfamily signalling seems to be at the biological basis of several loci, since genes such as *SMAD7*, *GREM1*, *BMP2*, *BMP4* and *RHPN2* are inside or near the CRC associating linkage-disequilibrium (LD) regions (13).

Although each of the ten variants independently confer only modest predisposition to CRC, their additive contribution on an individual's risk can be much higher. It has been estimated that individuals with 15 risk alleles in the ten loci are at 3.6-fold risk of developing CRC, compared with individuals possessing nine risk alleles (12). Clearly the individuals with many risk alleles can be predicted to display positive family history for CRC more often than others. However, ideal data sets – population-based material with full cancer information on family members – to examine the size of the effect are few. Recently, Middeldorp et al. (2009) studied six of the loci (at 8q24, 15q13, 8q23, 10p14, and 11q23) in a Dutch CRC cohort, concluding that the risk alleles were indeed enriched in the familial cases (16). The Dutch cohort consisted of 995 individuals that were selected based on early-onset of the disease or for having at least two affected first-degree relatives. Middeldorp et al. showed that the patients with a family history of CRC harboured significantly more risk alleles in the six loci than the early-onset solitary cases.

In this study, we have sought to establish the combined impact of the ten low-penetrance variants on the familial aggregation of CRC in an unbiased population-based series of 826 Finnish CRC cases, with verified cancer data from all first-degree relatives.

Materials and Methods

Patient samples

A previously characterized population-based sample series of 1042 CRCs collected between 1994 and 1998 from nine central hospitals in south-eastern Finland was utilized in this study (17, 18). As described in the previous studies, DNA was extracted from fresh-frozen normal tissue or blood with standard methods. Comprehensive registry-based clinical data and tumour characteristics evaluated by a pathologist were collected from the patients. Information on cancer occurrence in the first-degree relatives (parents, siblings, and offspring) of the 1,042 probands was obtained from the cancer registry, death certificates, and medical records. Microsatellite instability (MSI) status was determined for all the tumours, and subsequently the patients with MSI tumours were screened for mutations in *MLH1* and *MSH2* genes in order to identify Lynch syndrome patients. Diagnoses of other high-penetrance syndromes, including *MYH*-associated polyposis, familial adenomatous polyposis and juvenile polyposis, were also established based on clinical features and mutational analysis (3). Samples and clinicopathological information were obtained with informed consent and ethical review board approval in accordance with the tenets of the declaration of Helsinki.

Genotyping

All of the ten SNPs included in this study had previously been genotyped in the given sample series. Briefly, rs6983267, rs4779584 and rs4939827 were genotyped by direct genomic sequencing using Applied Biosystems BigDye v3.1 sequencing chemistry and ABI3730 Automatic DNA sequencer (12, 19). Competitive allele-specific PCR KASPar chemistry (KBiosciences Ltd, UK) was utilized for genotyping of rs3802842, rs16892766, rs10795668, rs4444235, rs9929218, rs10411210, and rs961253 (9, 11).

Apart from the 773 cases of which all of the ten SNPs had successfully been genotyped, there were 60 cases in which one of the genotypes was missing due to a failure in the analysis. The rerun was feasible for 53 samples that were sequenced for the missing SNP locus, resulting in 826 successfully genotyped individuals. The additional sequencing created no bias, since all of the samples with one missing genotype were included in the rerun, irrespective of how many risk alleles the individuals were known to harbour. Genomic DNA extracted from either fresh-frozen normal tissue or blood was amplified by standard PCR using AmpliTaq Gold DNA polymerase (Applied Biosystems, USA). Primers used in the PCR reactions were designed with Primer3 software, based on reference sequences from the Ensembl database. PCR reactions were purified with ExoSAP-IT enzyme (USB Corporation, USA) and sequencing was performed with Applied Biosystems BigDye v3.1 sequencing chemistry and ABI3730 Automatic DNA sequencer.

Statistical analysis

Odds ratios, 95% confidence intervals (95% CIs), and *P*-values for the associations of the ten SNPs with CRC were calculated with Pearson's Chi-squared test by comparing the allele frequencies between cases and controls. Information was combined from multiple SNPs by using an allele count model summing the number of risk alleles (0-20) carried by each individual. This assumes that each of the alleles has an equal and additive effect on CRC risk. The associations of these numbers with familial CRC were then analyzed using logistic regression. In the regression model, the relationship of the risk allele numbers as a

continuous predictor variable was analyzed with familial or sporadic CRC as a binary outcome variable. Allele groups below five were excluded from the analysis due to the small number of cases (two individuals with three risk alleles and four with four risk alleles). The effect of potential confounders, age at diagnosis and gender, was examined analyzing them as covariables with risk alleles. Adjustment for age at diagnosis improved model fit and was therefore included in analyses. This was categorized in ten year intervals (0-39, 40-49, 50-59, 60-69, 70-79, 80-) and considered as a factor variable in the model. The model fit was evaluated by comparing null deviance with residual deviance. The commonest risk allele number in the entire sample series was ten, which was used as a reference group (OR 1.00) when calculating odds ratios. The association of risk allele numbers with other familial cancers and several tumour characteristics (MSI status, Duke's stage, histological grade, and location) was also tested. Pearson's Chi-squared test with Yates' continuity correction was used to calculate potential differences in clinical and demographic characteristics between the cases with familial and sporadic CRC. R software (vs. 2.10.0) was used in all the above analyses.

Expected numbers of CRC cases in first-degree relatives of probands were computed in a previous study based on age-, sex-, and calendar period incidence rates for the general population of Finland (3). This was performed using the person-years program (20). Relatives were censored at the date of pedigree ascertainment, emigration or last contact with the proband. Cancer incidence in relatives was truncated at age 80 years to avoid the inherent problems of mis-certification of registered cancer diagnosis and death, which if discounted represents a source of bias. Using these data we computed the proportion of the phenotypic variance in excess familial risk associated with the ten risk loci by Poisson regression. These analyses were conducted using the statistical software program STATA (Version 10; Stata Corporation, Texas, USA).

Results

The tagging SNPs analyzed were rs6983267 at 8q24, rs4779584 at 15q13, rs4939827 at 18q21, rs16892766 at 8q23, rs10795668 at 10p14, rs3802842 at 11q23, rs4444235 at 14q22, rs9929218 at 16q22, rs10411210 at 19q13, and rs961253 at 20p12. Our sample series of ~1,000 CRC cases and population-matched controls was included in the replication phase of the UK GWA study and in the meta-analysis (9, 11, 12). Results from the genotyping studies of these Finnish samples are summarized in Table 1.

Altogether 826 individuals from a population-based CRC series had been or were successfully genotyped for all the ten variants. Out of the 826 cases, 37 were known to carry germline mutations (in either *MLH1*, *MSH2*, *MSH6*, *MYH*, *APC*, or *ALK3*) and were therefore excluded from the analysis. Out of the remaining 789 patients, 97 had a family history of CRC (at least one affected first-degree relative) and 691 were sporadic cases. One individual was excluded from the analysis due to non-Finnish descent and missing family history records. There were altogether 426 probands with other cancers than CRC in the first-degree relatives. Detailed demographic and clinicopathological characteristics of the 97 familial and 691 sporadic patients included in this study are reported in Table 2. Age at diagnosis in the familial cases was slightly higher than in the sporadic cases (69.8 years, and 67.8 years respectively; $P=0.08$). This could not be attributed to any effects of the low-penetrance variants, since the median number of risk alleles was ten in both younger (<69 years) and older cases (≥69 years). There were no significant difference between the familial and sporadic cases, with respect to gender ($P=0.74$), MSI status ($P=0.94$), Duke's stage ($P=0.25$), histological grade ($P=0.64$), or tumour location ($P=0.50$) (Table 2).

The overall number of risk alleles (1-20) in the ten SNPs was calculated for each of the 788 cases. Figure 1. shows the distribution of the number of risk alleles in familial and sporadic cases. The median number of risk alleles in familial cases was 11 compared with 10 for sporadic cases. There was a linear association between the number of risk alleles and familial CRC ($P_{\text{Trend}}=0.006$). For each one risk allele increase, the odds of having familial cancer increased by a factor of 1.16 (95% CI 1.04-1.30; Table 3). The increasing odds of having familial CRC were also described as odds ratios, comparing each of the risk allele groups with that of ten risk alleles, which was the most common number of risk alleles in the entire sample series. The individuals with 15 risk alleles were more than twice as likely to have affected first-degree relatives compared with the individuals possessing 10 risk alleles (Table 3).

The association of risk allele numbers with other familial cancers and several tumour characteristics, including MSI status, Duke's class, histological grade and location was examined. No relationship between the number of risk alleles and other familial cancers than CRC was found (OR=1.06, 95% CI 0.99-1.14, $P=0.10$). A trend towards the association of increasing numbers of risk alleles with distal or left-sided, rather than proximal or right-sided, location of the tumour was observed (OR=1.08, 95% CI 1.00-1.17, $P=0.05$), but not when based on comparison of colonic versus rectal disease ($P=0.23$). There was no association between the risk allele numbers and any other tumour characteristics: MSI status (OR=0.93, 95% CI 0.83-1.04, $P=0.23$), Duke's class (OR=1.03, 95% CI 0.96-1.11, $P=0.40$), or histological grade (OR=0.91, 95% CI 0.80-1.04, $P=0.18$).

Finally, the excess familial risk explained by the ten low-penetrance variants was calculated. This estimates which proportion of the familial aggregation of CRC in our sample series can be attributed to the given loci. Altogether 84 first-degree relatives were diagnosed with CRC by age 80 compared with 60.3 expected. On the basis of poisson regression analysis of the observed CRC in first-degree relatives accounting for expected numbers, the ten loci were calculated to be responsible for 8.7% of the total variance in familial CRC risk (upper 95% CI= 19.0%).

Discussion

Excluding the high-penetrance CRC syndromes, around 10% of the CRC patients have a family history of the disease (3). This non-syndromic familial CRC can be defined as an apparently sporadic form of the disease that occurs in families more often than expected by chance. Several common low-penetrance variants are likely to contribute to this predisposition and their identification may enable genetic risk profiling and tailoring of preventive interventions. Although empirical risk stratification based on family history of CRC is already feasible, it does not fully address differences in the inheritance of risk alleles among offspring and hence different risk profiles.

In this study, we analyzed the contribution of all the recently identified ten low-penetrance CRC variants to the familial aggregation in a Finnish population-based sample series. Results of the GWA study's replication in the Finnish sample series show that most of the ten associations are well replicated in our sample series.

In our study it is striking that the impact of the ten variants on the inherited predisposition to CRC is apparent despite its modest size. We observed a clear association of increasing numbers of risk alleles in the ten loci with familial CRC, compatible with the observation of Middeldorp et al. (2009) from six of the loci (16). The 995 Dutch cases were selected based on family history or early-onset of CRC, whereas our sample series was population-based.

Middeldorp et al. also reported an association of risk allele number with early-onset, compared with late-onset disease, which was not observed in our study.

Three low-penetrance loci have been reported to be more common in rectal than in colonic tumours: rs3802842 at 11q23, rs4939827 at 18q21, and rs10795668 at 10p14 (9, 10). We also saw a trend towards the association of increasing numbers of risk alleles and odds of having distal rather than proximal cancer ($P=0.04$). No enrichment of risk alleles was detected in any subgroup in terms of age at diagnosis, gender, Duke's stage, histological grade, or MSI status of the tumour. This suggests that in combination, the ten low-penetrance variants do not clearly contribute to a specific type of CRC, but have a generic influence. We did not find any association between increasing numbers of risk alleles in the ten loci and odds of having other familial cancer. This was expected, since pleiotropic effects are reported only in the 8q24 CRC locus that increases the risk of prostate and ovarian cancers (6, 21).

In a meta-analysis of the UK GWA studies, the excess familial risk of CRC attributable to the ten low-penetrance variants was calculated to be 6% (12). This was, however, an indirect estimate based on a log-additive model. To address the contribution of the ten loci to the familial risk of CRC we directly calculated the excess familial risk using incidence data from the first-degree relatives of the genotyped probands. While a limitation of this study is its relatively small sample size, this is the first time the excess risk associated with the ten variants has been estimated using comprehensive registry-based family histories in a well-characterized population-based sample series. This is a major advantage, since self-reported family histories tend to be inaccurate. Furthermore, we used age-, sex-, and calendar period incidence rates for the general population of Finland for comparison when calculating the expected incidences in the first-degree relatives of the genotyped probands, which minimizes bias. Our observation that ~9% of the variance in familial risk is explained by the ten loci provides evidence to support the indirect estimations that the majority of the inherited predisposition to CRC remains unaccounted for by the currently known variants.

As new low-penetrance variants are identified in even larger GWA studies and meta-analyses, their role in explaining the familial risk of CRC is likely to be better refined. Tenesa and Dunlop (2009) have estimated that up to ~170 common variants could independently contribute to the observed hereditary predisposition to CRC (13). On the other hand, the existing GWA studies have captured the majority of the common variation in European populations and hence, finding of high numbers of additional low-penetrance variants seems unlikely. In order to facilitate the characterization of CRC predisposition, an international consortium, COGENT (COlorectal cancer GENEtics) has been established that currently has access to over 48,000 cases and 43,000 controls from 20 different research groups all over the world (22).

An important part of the familial risk might also be explained by moderate-penetrance loci, as these have not been well captured by the current genotyping platforms. Next-generation sequencing efforts are expected to play a crucial role in discovering these variants. Hemminki et al. (2008) have suggested that many of the low-penetrance associations could actually be markers of rarer functional alleles and hence explain a larger part of the excess familial risk than currently estimated (23). However, this is not supported by the causal variant in 8q24 being the tagging SNP itself, or by the 18q21 locus where the causal variant has a frequency of 43% and confers a similar increase in risk ($OR=1.22$) than the tagging SNP (14, 15). It remains to be established whether the functional changes in any of the remaining eight loci are moderate-penetrance variants.

In summary, by statistically analyzing the currently known ten low-penetrance CRC variants in Finland, we have shown a significant association between increasing numbers of risk alleles in the ten loci and odds of having familial, rather than sporadic CRC. Using registry-based family history data, we estimated the ten loci to underlie ~9% of the variance in familial risk. This study contributes to the understanding of the genetic landscape of CRC but also emphasizes the need to characterize the remaining inherited predisposition to this malignancy.

Acknowledgments

We thank Sini Marttinen for managing the patient data.

Grant support: This work was supported by the Academy of Finland (Centre of Excellence in Translational Genome-Scale Biology grant 6302352), the Finnish Cancer Society, and the Sigrid Juselius Foundation; by grants to I.N. from the Paulo Foundation and the Finnish Cancer Society; and by grants to R.S.H. from Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund).

References

1. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol.* 2001; 96:2992–3003. [PubMed: 11693338]
2. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer - analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000; 343:78–85. [PubMed: 10891514]
3. Aaltonen L, Johns L, Jarvinen H, Mecklin JP, Houlston R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin Cancer Res.* 2007; 13:356–61. [PubMed: 17200375]
4. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet.* 2007; 39:984–8. [PubMed: 17618284]
5. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2007; 39:989–94. [PubMed: 17618283]
6. Haiman CA, Le Marchand L, Yamamoto J, et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet.* 2007; 39:954–6. [PubMed: 17618282]
7. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet.* 2007; 39:1315–7. [PubMed: 17934461]
8. Jaeger E, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet.* 2008; 40:26–8. [PubMed: 18084292]
9. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet.* 2008; 40:623–30. [PubMed: 18372905]
10. Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet.* 2008; 40:631–7. [PubMed: 18372901]
11. Pittman AM, Webb E, Carvajal-Carmona L, et al. Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer. *Hum Mol Genet.* 2008; 17:3720–7. [PubMed: 18753146]
12. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet.* 2008; 40:1426–35. [PubMed: 19011631]
13. Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat Rev Genet.* 2009; 10:353–8. [PubMed: 19434079]

14. Tuupanen S, Turunen M, Lehtonen R, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced wnt signaling. *Nat Genet.* 2009; 41:885–90. [PubMed: 19561604]
15. Pittman AM, Naranjo S, Webb E, et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res.* 2009; 19:987–93. [PubMed: 19395656]
16. Middeldorp A, Jagmohan-Changur S, van Eijk R, et al. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. *Cancer Epidemiol Biomarkers Prev.* 2009; 18:3062–7. [PubMed: 19843678]
17. Aaltonen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med.* 1998; 338:1481–7. [PubMed: 9593786]
18. Salovaara R, Loukola A, Kristo P, et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J Clin Oncol.* 2000; 18:2193–200. [PubMed: 10829038]
19. Tuupanen S, Niittymäki I, Nousiainen K, et al. Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. *Cancer Res.* 2008; 68:14–7. [PubMed: 18172290]
20. Coleman M, Douglas A, Hermon C, Peto J. Cohort study analysis with a FORTRAN computer program. *Int J Epidemiol.* 1986; 15:134–7. [PubMed: 3754239]
21. Ghossaini M, Song H, Koessler T, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst.* 2008; 100:962–6. [PubMed: 18577746]
22. Tomlinson IP, Dunlop M, Campbell H, et al. COGENT (COlorectal cancer GENEtics): An international consortium to study the role of polymorphic variation on the risk of colorectal cancer. *Br J Cancer.* 2010; 102:447–54. [PubMed: 19920828]
23. Hemminki K, Forsti A, Bermejo JL. The ‘common disease-common variant’ hypothesis and familial risks. *PLoS One.* 2008; 3:e2504. [PubMed: 18560565]

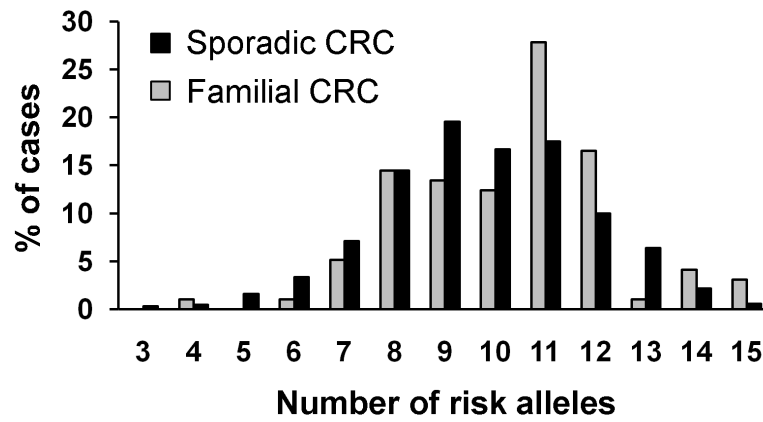


Figure 1.

Distribution of risk alleles in ten low-penetrance variants among familial (n=97) and sporadic (n=691) CRC. In familial cases, the most common risk allele number was 11 (27/97 cases, 27.8%) and in sporadic cases 9 (135/691 cases, 19.5%). There were no cases with less than three or more than 15 risk alleles.

Table 1

Summary of the low-penetrance CRC variants in the Finnish population

| SNP ID | Chr. location | Risk allele | Risk allele frequency [§] | Cases | Controls | Allelic OR* (95% CI) | P | Study |
|------------|---------------|-------------|------------------------------------|-------|----------|----------------------|-------|-----------------------------|
| rs6983267 | 8q24 | G | 0.53 | 996 | 1012 | 1.22 (1.08-1.38) | 0.002 | Tuupanen <i>et al.</i> (19) |
| rs4779584 | 15q13 | T | 0.3 | 981 | 1024 | 1.14 (1.00-1.31) | 0.053 | Houlston <i>et al.</i> (12) |
| rs4939827 | 18q21 | T | 0.46 | 970 | 969 | 1.20 (1.06-1.36) | 0.004 | Houlston <i>et al.</i> (12) |
| rs16892766 | 8q23 | C | 0.12 | 968 | 1003 | 1.21 (1.01-1.46) | 0.041 | Tomlinson <i>et al.</i> (9) |
| rs10795668 | 10p14 | G | 0.69 | 951 | 983 | 1.19 (1.04-1.37) | 0.014 | Tomlinson <i>et al.</i> (9) |
| rs3802842 | 11q23 | C | 0.24 | 970 | 994 | 1.21 (1.05-1.40) | 0.008 | Pittman <i>et al.</i> (11) |
| rs4444235 | 14q22 | C | 0.43 | 933 | 828 | 1.16 (1.02-1.33) | 0.028 | Houlston <i>et al.</i> (12) |
| rs9929218 | 16q22 | G | 0.77 | 945 | 836 | 1.03 (0.88-1.20) | 0.751 | Houlston <i>et al.</i> (12) |
| rs10411210 | 19q13 | C | 0.84 | 917 | 795 | 1.18 (0.97-1.42) | 0.091 | Houlston <i>et al.</i> (12) |
| rs961253 | 20p12 | A | 0.29 | 941 | 814 | 1.23 (1.07-1.42) | 0.005 | Houlston <i>et al.</i> (12) |

[§]In our controls,

*Risk allele versus neutral allele

Table 2

Characteristics of 788 analyzed colorectal cancer cases

| Characteristic | Familial (n=97) | Sporadic (n=691) | <i>P</i> |
|--------------------|--------------------|---------------------|----------|
| Age at diagnosis | | | |
| 69 | 59 | 351 | 0.08 |
| <69 | 38 | 340 | |
| Gender | | | |
| M | 52 | 354 | 0.74 |
| F | 45 | 337 | |
| MSI status | | | |
| MSI | 11 | 76 | 0.94 |
| MSS | 86 | 615 | |
| Duke's stage | | | |
| A-B | 62 | 398 | 0.25 |
| C-D | 34 | 291 | |
| Histological grade | | | |
| I-II | 84 | 625 | 0.64 |
| III-IV | 9 | 52 | |
| Location * | | | |
| Distal | 62 | 476 | 0.5 |
| Proximal | 33 | 211 | |

Some of the numbers do not match do to missing data.

* Distal = from splenic flexure to rectum, proximal = from cecum to transverse colon

Table 3

Odds ratios for familial cancer by number of risk alleles

| Number of risk alleles | Familial cases n(%) | Sporadic cases n(%) | Odds ^A | OR if 10 ref.* |
|------------------------|---------------------|---------------------|--|----------------|
| 15 | 3 (3.1) | 4 (0.6) | 0.2 | 2.15 |
| 14 | 4 (4.1) | 15 (2.2) | 0.18 | 1.84 |
| 13 | 1 (1.0) | 44 (6.4) | 0.15 | 1.58 |
| 12 | 16 (16.5) | 69 (10.0) | 0.13 | 1.36 |
| 11 | 27 (27.8) | 121 (17.5) | 0.11 | 1.17 |
| 10 | 12 (12.4) | 115 (16.6) | 0.1 | 1 |
| 9 | 13 (13.4) | 135 (19.5) | 0.08 | 0.86 |
| 8 | 14 (14.4) | 100 (14.5) | 0.07 | 0.74 |
| 7 | 5 (5.2) | 49 (7.1) | 0.06 | 0.63 |
| 6 | 1 (1.0) | 23 (3.3) | 0.05 | 0.54 |
| 5 | 0 | 11 (1.6) | 0.04 | 0.47 |
| Total | 97 | 691 | OR=1.16 (1.04-1.30) per allele $P_{\text{Trend}}=0.006$ | |

^A=Adjusted with age at diagnosis

*Odds ratio if 10 risk alleles (the most common number in all cases) set as a reference group