Quantifying the utility of single nucleotide polymorphisms to guide colorectal cancer screening

Mark A Jenkins*[,1], Enes Makalic[1], James G Dowty[1], Daniel F Schmidt[1], Gillian S Dite[1], Robert J MacInnis[1,2], Driss Ait Ouakrim[1], Mark Clendenning[3], Louisa B Flander[1], Oliver K Stanesby[1], John L Hopper[1], Aung K Win[1] & Daniel D Buchanan[1,3]

Aim: To determine whether single nucleotide polymorphisms (SNPs) can be used to identify people who should be screened for colorectal cancer. **Methods:** We simulated one million people with and without colorectal cancer based on published SNP allele frequencies and strengths of colorectal cancer association. We estimated 5-year risks of colorectal cancer by number of risk alleles. **Results:** We identified 45 SNPs with an average 1.14-fold increase colorectal cancer risk per allele (range: 1.05–1.53). The colorectal cancer risk for people in the highest quintile of risk alleles was 1.81-times that for the average person. **Conclusion:** We have quantified the extent to which known susceptibility SNPs can stratify the population into clinically useful colorectal cancer risk categories.

Background & aim

Genetic susceptibility to inherited colorectal cancer is complex and involves multiple variants and genes. For example, several genes (such as the DNA mismatch repair genes, *MUTYH* and *APC*), if inherited in a mutated form, place the carrier at a high risk of colorectal cancer [1]. Although mutations in these genes are very rare, cancer risk is sufficiently high to warrant guidelines on who to test for mutations in these genes, and who to screen for cancer, in order to reduce the burden of cancer for mutation carriers. There are also an increasing number of known common genetic variants, known as single nucleotide polymorphisms (SNPs), which are susceptibility markers with each associated with a small increased risk of colorectal cancer [2]. These have been discovered from genome-wide association studies that have used germline DNA extracted from blood samples to measure hundreds of thousands, and even millions, of SNPs across the genome and have compared the frequency of alleles of these SNPs in people with a specific cancer with those without that cancer [3]. Why these SNPs, many of which do not appear to be within the coding regions of genes, are associated with cancer risks is largely unknown [4].

In the context of cancer risk estimation, it is not necessary to know the causes of why the SNPs are associated with disease risk (although knowing the true causal variants responsible for these associations would of course be preferable). Instead, it is more important to know whether the SNPs can be used to stratify people in the population in terms of their risk of developing cancer. While each SNP risk allele is associated with only a small increase in risk (e.g., 5–25%) [5–7], carriers of many

[1]Centre for Epidemiology & Biostatistics, Melbourne School of Population & Global Health, The University of Melbourne, Parkville Victoria, VIC 3010, Australia
[2]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC 3004, Australia
[3]Colorectal Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, School of Medicine, The University of Melbourne, Parkville Victoria, VIC 3010, Australia
*Author for correspondence: Tel.: +61 383 440 902; Fax: +61 393 495 815; m.jenkins@unimelb.edu.au

SNP risk alleles can have several times the cancer risk of those with only a few SNP risk alleles. In 2012, the cumulative impact on colorectal cancer risk for ten of these SNPs was assessed, with investigators concluding that people in the upper tail of the distribution for number of risk alleles reached a threshold of risk for which colonoscopy, rather than the less invasive fecal occult blood testing, would be recommended [8]. Since this report, more independent SNPs associated with colorectal cancer risk have been identified as a result of intensive research and pooling of data from case–control studies.

Our aim was to quantify the utility of the SNPs currently known to be associated with colorectal cancer risk as a tool to stratify people by personal risk of colorectal cancer so as to help make decisions about colorectal cancer screening regimens appropriate for that risk.

Methods

We conducted a literature review to identify all publications reporting associations between SNPs and colorectal cancer. To do this we used PubMed using the search terms 'genome-wide association study', 'GWAS', 'colorectal cancer', 'colon cancer', 'SNP' and 'single nucleotide polymorphism'. We then searched the references of each identified publication to identify any additional publications of SNPs. We restricted publications to those that either reported internal validation of identified SNP associations with colorectal cancer, or were independent publications confirming previously reported SNP colorectal cancer associations. We also restricted publications to those reporting studies of people of European descent.

Because our aim was to examine the independent contribution of risk alleles to colorectal cancer risk, we excluded SNPs for which there was sufficient evidence that their association with colorectal cancer was due to linkage disequilibrium with another SNP within the same region. Primary evidence for linkage disequilibrium was a lack of a statistically significant association when tested in a multivariable logistic regression model with other SNPs within the region. Where no such analyses were conducted, secondary evidence for linkage disequilibrium was a D prime with the other SNPs within the region >0.5. Where linkage disequilibrium was identified, we excluded the SNP that had the least consistent association with colorectal cancer, was from the smallest study or had the weakest association with

colorectal cancer (considered in this order). For each reported SNP, we recorded the allele frequency of the risk allele for colorectal cancer, in other words, the allele reported to be associated with an increased risk of colorectal cancer; and the odds ratio per risk allele.

Using the software PLINK [9,10], we conducted a simulation to determine the ability of the cumulative number of risk alleles of the SNPs to discriminate cases of colorectal cancer from controls and to estimate the risk of colorectal cancer as a function of the number of risk alleles. We simulated a population of 1,000,000 people with colorectal cancer (cases) and 1,000,000 people without colorectal cancer (controls). The distribution of SNP risk alleles for the simulated population was matched to the reported risk allele frequencies and per allele odds ratios of colorectal cancer associations. We assumed a simplistic model of risk where the association with colorectal cancer for each SNP was independent. This is consistent with a previous analysis on a subset of these SNPs [11] which found their associations across SNPs were additive on a log scale. In this analysis we also assumed that the odds ratios reported for colorectal cancer for each SNP were applicable to both men and women and were constant with age.

We assessed the discriminatory power of the SNPs to distinguish cases from controls using a receiver operating curve and estimating the AUC (the probability that a randomly selected colorectal cancer case will have more risk alleles than a randomly selected control). We estimated the odds ratios for colorectal cancer risk for: being in the highest and lowest quintile for the number of risk alleles being in the middle quintile; being in the highest and lowest decile for the number of risk alleles versus being in the median number of risk alleles; and per standard deviation of number of risk alleles. Cutoffs for number of risk alleles for quintiles and deciles, and the standard deviation, were based on the distribution of risk alleles for the controls.

Under the assumption that these odds ratios were constant with age and equal for men and women, we estimated the cumulative lifetime risk (from birth to age 70 years) and the 5-year risk for each age category of colorectal cancer for men and for women in Australia and the USA by the number of SNP risk alleles. We assumed that the age-specific Australian and US population incidences were the incidences for those with the median number of risk alleles. Colorectal cancer population incidences were obtained from the

Australian Institute of Health and Welfare [12] and the Surveillance, Epidemiology, and End Results (SEER) Program Cancer Statistics [13].

We estimated the proportion of log familial relative risk (FRR; the odds ratio for colorectal cancer associated with having a first-degree relative with colorectal cancer) that could be attributable to the risk alleles of the SNPs. We assumed Hardy–Weinberg equilibrium for each SNP, linkage equilibrium between the SNPs and a multiplicative model for the associations of the SNPs with colorectal cancer risk. More precisely, let SNP_1, ., SNP_{45} be the known colorectal cancer-associated SNPs and let SNP_{46}, ., SNP_m be unknown ones (Note: SNP_{46}, ., SNP_m these could be any heritable factors contributing to the FRR, but for simplicity we think of them as SNPs). Then if G_i is a random variable giving the number of risk alleles at SNP_i for a random person from the population, then G_1, ., G_m, ., are all independent random variables (by linkage equilibrium) and the log-odds ratio for a random person is $X_1 + \ldots + X_m$ (by the assumed multiplicative model), where $X_i = G_i \log OR_i$ and OR_i is the per-allele odds ratio for SNP_i. A formula of Antoniou *et al.* [14], derived rigorously in Win *et al.* [15], then becomes

$$\log FRR = \tfrac{1}{2}[Var(X_1) + \cdots + Var(X_m)]$$

.

This shows that the log FRR is the sum of independent components from the known and unknown colorectal cancer-associated SNPs. The proportion of the log FRR due to the known SNPs is

$$\tfrac{1}{2}(Var(X_1) + \cdots + Var(X_{45}))/\log FRR,$$

while the proportion due to the unknown SNPs is one minus this value. We assumed that the FRR of having at least one first-degree relative with colorectal cancer was 2.25, based on a previous meta-analysis of family history of colorectal cancer [16], and an elementary calculation (assuming Hardy–Weinberg equilibrium) shows that

$$Var(X_i) = 2\,p_i(1 - p_i)(\log OR_i)^2$$, where p_i

is the minor allele frequency of SNP_i. Using this statistic, we estimated the 5-year risk of colorectal cancer by the number of risk alleles, with and without a family history of colorectal cancer.

Results

We identified 55 SNPs within 39 regions reported to be associated with colorectal cancer in European populations. Of these, four SNPs within 11q12.2

(rs174537, rs4246215, rs174550 and rs1535) were reported to be perfectly correlated and could be represented by a common haplotype [17] (named here as the 11q12.2 haplotype). Two SNPs within 19q13.2 (rs1800469 and rs2241714) were reported to be perfectly correlated and could be represented by a common haplotype [17] (named here as the 19q13.2 haplotype). One SNP is on the X chromosome (rs5934683) [18] and was not included in our simulation of colorectal cancer risk for males and females combined. Two SNPs within 1q41 (rs6687758 and rs6691170) were shown to be in linkage disequilibrium with a single imputed SNP [19], and we therefore excluded rs6691170. Three SNPs within 8q24.21 (rs10505477, rs6983267 and rs7014346) had a D prime of 1.0, and based on the results of a meta-analysis [20], we excluded rs10505477 and rs7014346. Three of the SNPs are within 15q31 (rs11632715, rs16969681 and rs4779584), but because only two were associated with colorectal cancer in a multivariable model [21], we excluded rs4779584. Two of the SNPs are within 10q24.2 (rs1035209, rs11190164) with a D prime of 0.9, therefore, we excluded rs1035209. Two SNPs are within 12q13.13 (rs11169552, rs7136702) with a D prime of 0.6, however, as they were shown to be independently associated with colorectal cancer [19], both SNPs were included. All other co-located SNPs (rs3217810, rs3217901 and rs10774214 within 12p13.32; rs1957636 and rs4444235 within 14q22.2 and rs2423279, rs4813802 and rs961253 within 20p12.3) had a D prime of <0.5 and were retained. Details of the remaining 45 independent autosomal SNPs are provided in **Table 1**.

The average risk allele frequency was 0.43 (range: 0.07–0.91). The average odds ratio per risk allele was 1.14 (range: 1.05–1.53). The average FRR that could be attributed to each SNP was 1.0040 (range: 1.0006–1.0281), which is 0.50% (range: 0.07–3.41%) of the total log FRR. The combined FRR that could be attributable to all 45 SNPs was 1.1980, which is 22.3% of the total log FRR. The estimated FRR not due to the SNPs was 1.88.

There was considerable overlap for the number of risk alleles for the simulated people with and without colorectal cancer (for those with colorectal cancer: median 42 risk alleles, range 21–61 risk alleles, mean 41.6 risk alleles, standard deviation 4.2 risk alleles; for those without colorectal cancer: median 40 risk alleles, range 20–59, mean 39.7 risk alleles, standard deviation 4.2 risk alleles; upper quartile 44 or more

Table 1. Single nucleotide polymorphisms reported to be associated with colorectal cancer (independent of other single nucleotide polymorphisms) in European populations.

Locus	Gene†	SNP	Per risk allele OR	Frequency of risk allele	FRR	Proportion of log FRR	Ref.
1p36.2	*WNT4; CDC42*	rs72647484	1.21	0.91	1.003	0.37%	[22]
1q25.3	*LAMC1*	rs10911251	1.05	0.54	1.0006	0.07%	[23,24]
1q41	*DUSP10; CICP13*	rs6687758	1.09	0.2	1.0012	0.15%	[25]
2q32.3	*NABP1; MYO1B; SDPR*	rs11903757	1.06	0.36	1.003	0.37%	[25,26]
3p14.1	*LRIG1*	rs812481	1.09	0.58	1.0018	0.22%	[27]
3p22.1	*RP11; CTNNB1*	rs35360328	1.14	0.16	1.0023	0.29%	[27]
3q26.2	*MYNN; TERC*	rs10936599	1.08	0.75	1.0011	0.14%	[25]
4q26	*NDST3*	rs3987	1.36	0.44	1.0235	2.87%	[28]
4q32.2	*FSTL5*	rs35509282	1.53	0.09	1.0149	1.83%	[29]
5q31.1	*PITX1; H2AFY*	rs647161	1.11	0.67	1.0024	0.30%	[30]
6p21.31	*CDKN1A*	rs1321311	1.1	0.23	1.0016	0.20%	[18]
8q23.3	*EIF3H*	rs16892766	1.25	0.07	1.0032	0.40%	[31]
8q24.21	*CCAT2; MYC*	rs6983267	1.21	0.52	1.0091	1.12%	[32,33]
9q24	*TPD52L3; UHRF2*	rs719725	1.19	0.37	1.0011	0.13%	[32,34]
10p13	*CUBN*	rs10904849	1.14	0.68	1.0037	0.46%	[22]
10p14	*GATA3*	rs10795668	1.12	0.67	1.0028	0.35%	[31]
10q22.3	*ZMIZ1; AS1*	rs704017	1.06	0.57	1.0008	0.10%	[17]
10q24.2	*SLC25A28; ENTPD7; COX15; CUTC; ABCC2*	rs11190164	1.09	0.29	1.0015	0.19%	[27]
10q25	*VTI1A*	rs12241008	1.13	0.09	1.0012	0.15%	[17]
11q12.2	*FADS1; FEN1*	11qhap‡	1.4	0.57	1.0281	3.41%	[17]
11q13.4	*POLD3*	rs3824999	1.08	0.5	1.0015	0.18%	[18]
11q23.1	*COLCA2*	rs3802842	1.11	0.29	1.0022	0.28%	[35]
12p13.32	*CCND2*	rs3217810	1.2	0.16	1.0045	0.55%	[23,24]
12p13.32	*CCND2*	rs3217901	1.1	0.41	1.0022	0.27%	[23,24]
12p13.32	*CCND2*	rs10774214	1.09	0.38	1.0018	0.22%	[30]
12q13.13	*DIP2B; ATF1*	rs11169552	1.09	0.72	1.0015	0.18%	[25]
12q13.13	*LARP4; DIP2B*	rs7136702	1.06	0.35	1.0008	0.10%	[25]
12q24.12	*SH2B3*	rs3184504	1.09	0.53	1.0019	0.23%	[27]
12q24.21	*TBX3*	rs59336	1.09	0.48	1.0019	0.23%	[26]
12q24.22	*NOS1*	rs73208120	1.16	0.11	1.0021	0.26%	[27]
14q22.2	*BMP4*	rs1957636	1.08	0.4	1.0014	0.18%	[36]
14q22.2	*BMP4*	rs4444235	1.11	0.46	1.0027	0.33%	[36,37]
15q13.3	*SCG5; GREM1*	rs11632715	1.12	0.47	1.0032	0.39%	[36]
15q13.3	*SCG5; GREM1*	rs16969681	1.18	0.09	1.0022	0.28%	[36]
16q22.1	*CDH1*	rs9929218	1.1	0.71	1.0019	0.23%	[37]
16q24.1	*FOXL1*	rs16941835	1.15	0.21	1.0032	0.40%	[22]
17q21	*STAT3*	rs744166	1.27	0.55	1.0142	1.74%	[38]
18q21.1	*SMAD7*	rs4939827	1.18	0.52	1.0069	0.84%	[35,39]
19q13.11	*RHPN2*	rs10411210	1.15	0.9	1.0018	0.22%	[37]
19q13.2	*TMEM91; TGFB1*	19qhap‡	1.16	0.49	1.0055	0.68%	[17]
20p12.3	*FERMT1; BMP2*	rs2423279,	1.14	0.3	1.0036	0.44%	[24,30]
20p12.3	*FERMT1; BMP2*	rs4813802	1.09	0.36	1.0017	0.21%	[36],

SNPs reported to be associated with colorectal cancer (independent of other SNPs) in European populations, including the SNP nomenclature, the gene(s) closest to or within the likely regulatory target of the SNPs, the reported risk allele genotype, the reported risk allele frequency in controls, the reported association with colorectal cancer per risk allele (odds ratio), the FRR attributable to the SNPs and the proportion of the log FRR due to the SNPs.
†Gene/s closest to or likely regulatory target of SNP.
‡Four SNPs within 11q12.2 (rs174537, rs4246215, rs174550, and rs1535) were reported to be perfectly correlated and could be represented by a common haplotype [17] (named here as the 11q12.2 haplotype). Two SNPs within 19q13.2 (rs1800469 and rs2241714) were reported to be perfectly correlated and could be represented by a common haplotype [17] (named here as the 19q13.2 haplotype).
FRR: Familial relative risk; SNP: Single nucleotide polymorphism.

Locus	Gene[†]	SNP	Per risk allele OR	Frequency of risk allele	FRR	Proportion of log FRR	Ref.
20p12.3	*FERMT1; BMP2*	rs961253	1.12	0.36	1.003	0.36%	[36,37]
20q13.1	*PREX1*	rs6066825	1.09	0.64	1.0017	0.21%	[27]
20q13.33	*LAMA5*	rs4925386	1.08	0.68	1.0013	0.16%	[25]

Table 1. Single nucleotide polymorphisms reported to be associated with colorectal cancer (independent of other single nucleotide polymorphisms) in European populations.

SNPs reported to be associated with colorectal cancer (independent of other SNPs) in European populations, including the SNP nomenclature, the gene(s) closest to or within the likely regulatory target of the SNPs, the reported risk allele genotype, the reported risk allele frequency in controls, the reported association with colorectal cancer per risk allele (odds ratio), the FRR attributable to the SNPs and the proportion of the log FRR due to the SNPs.
[†]Gene/s closest to or likely regulatory target of SNP.
[‡]Four SNPs within 11q12.2 (rs174537, rs4246215, rs174550, and rs1535) were reported to be perfectly correlated and could be represented by a common haplotype [17] (named here as the 11q12.2 haplotype). Two SNPs within 19q13.2 (rs1800469 and rs2241714) were reported to be perfectly correlated and could be represented by a common haplotype [17] (named here as the 19q13.2 haplotype).
FRR: Familial relative risk; SNP: Single nucleotide polymorphism.

risk alleles; lower quartile 36 or fewer risk alleles; upper decile 46 or more risk alleles; lower decile 34 or fewer risk alleles) **(Figure 1)**. Having 29 risk alleles corresponded to a lifetime risk of colorectal cancer of 1.4% for a person from Australia and 1.0% for a person from the USA. The respective risks for 36 risk alleles were 2.9 and 2.0%; for 43 risk alleles were 6.1 and 4.3%; and for 50 risk alleles were 12.5 and 8.8% **(Figure 1)**. Compared with people in the middle quintile for the number of risk alleles, the odds ratio for colorectal cancer was 1.81 for people in the highest quintile of number of risk alleles, and 0.51 for people in the lowest quintile; this is equivalent to a 3.55-fold inter-quintile risk (highest vs lowest quintile). Compared with people with the median of 40 risk alleles, the odds ratio for colorectal cancer was 2.27 for people in the highest decile of the number of risk alleles, and 0.45 for people in the lowest decile; this is equivalent to a 5.04-fold inter-decile risk (highest vs lowest decile). The odds ratio per standard deviation of risk alleles was 1.57. The receiver operating characteristic curve had an AUC of 0.63.

Based on the 2011 population incidence rates for colorectal cancer in Australia, the average cumulative risk of colorectal cancer to age 70 years was 3.3%. For people in the highest quintile for number of risk alleles, the cumulative risk was 5.9% (11.5% if they also had a first-degree relative with colorectal cancer and 5.5% if they did not) compared with 1.7% for people in the lowest quintile for number of risk alleles (3.2% if they also had a first-degree relative with colorectal cancer and 1.6% if they did not). For people in the highest decile for number of risk alleles, the cumulative risk was 7.4% (13.4% if they also had a first-degree relative with colorectal cancer and 6.9% if they did not) compared with 1.5% for people in the lowest decile for

number of risk alleles (2.8% if they also had a first-degree relative with colorectal cancer and 1.4% if they did not; **Figure 2A & B**). The estimates for males were on average approximately 13% higher and for females the estimates were on average 16% lower than for males and females combined **(Supplementary Figures 1 & 2)**.

The 5-year risk of colorectal cancer for the average (previously unaffected) person in Australia reaches 1% at the age of 63 years. The same 1% 5-year risk is attained approximately 7 years earlier for people in the highest quintile for number of risk alleles (and approximately 14 years earlier if they also had a family history of colorectal cancer) and approximately 10 years earlier for people in the highest decile for number of risk alleles (16 years earlier if they also had a family history; **Figure 2C & D & Table 2**). On average, males reached the 1% risk threshold 1–2 years earlier, and females reached the threshold on average 3–4 years later than for males and females combined **(Table 2)**.

Given that the population incidence rates of colorectal cancer in the USA are lower (particularly after age 50 years compared with Australia), the associated risks based on the number of risk alleles and family history are also lower than those for Australia **(Figure 3A & B, Supplementary Figures 3 & 4)**. In comparison, the same 1% risk is attained approximately 9 years earlier for people in the highest quintile for number of risk alleles (20 years earlier if they also had a family history of colorectal cancer) and approximately 12 years earlier for people in the highest decile for number of risk alleles (22 years earlier if they also had a family history; **Figure 3C & D & Table 2**). On average, males reached the 1% risk threshold 3–5 years earlier, and females reached the threshold on average 1–3 years later than for males and females combined **(Table 2)**.
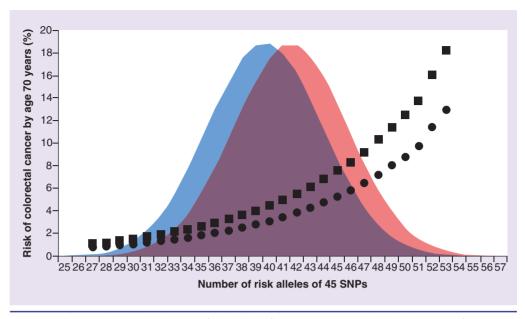
Figure 1. The simulated distribution of risk alleles for 1,000,000 people with a history of colorectal cancer (red) and 1,000,000 people without a history of colorectal cancer (blue); and the cumulative risk of colorectal cancer to age 70 years for the number of risk alleles for an Australian (square) and USA (circle) population.
SNP: Single nucleotide polymorphism.

Discussion

We have used simulations to quantify the utility of a panel of 45 risk-associated SNPs to categorize people based on their risk of colorectal cancer. Overall, the predictive power of the risk alleles for the 45 SNPs was not high; on average people with colorectal cancer had only two more risk alleles than those without colorectal cancer. However, people at the ends of the spectrum were considerably more likely to develop colorectal cancer (high end) or less likely to develop colorectal cancer (low end). Because the total variation in risk associated with these SNPs across the population can explain about one quarter of the total FRR, the predictive strength of the SNP profile is increased if family history of colorectal cancer is also taken into account. Given that the strength of association with colorectal cancer for those in the lowest 20% of the population (for number of risk alleles of these SNPs) is roughly the inverse of the increased risk associated with the remaining FRR, people who have a family history of colorectal cancer but who also are in the lowest quintile of the population for number of risk alleles of these SNPs, are at population risk.

Measurement of these SNPs is a potentially useful method for assessment of colorectal cancer risk, and could be a potential tool for determining who should be recommended for colorectal cancer screening, and at what intensity. For example, a person in the top 20% of the population for risk alleles (at least 44 alleles) reaches the average population 5-year risk 9 years earlier than the average person. Therefore, if the average person meets the risk threshold for fecal occult blood test screening (which most national screening programs recommend) at the age of 50 years, then it could be argued that a person with at least 44 risk alleles reaches the same risk threshold at the age of 41 years. The ages to begin colonoscopy screening for people with a first-degree relative with colorectal cancer would be 49 and 47 years for the highest quintile and the highest decile of risk alleles, respectively. In the USA, where the population risk of colorectal cancer is lower than for Australia, the 2% threshold for being in the top quintile or decile and having a family history of colorectal cancer is reached at ages 62 and 59 years, respectively.

Research is continually identifying new SNPs associated with colorectal cancer, and better risk scores based on SNPs are likely to be developed. We are aware of three comparable studies of Caucasian populations. The first was the study by Dunlop *et al.* [8] of ten SNPs within ten regions;

the second was the study by Lubbe *et al.* [11] of 14 SNPs within 14 regions (including SNPs within all of the regions covered by Dunlop *et al.* plus an additional four SNPs within four regions); the third was the study by Hsu *et al.* [40] of 27 SNPS within 22 regions (including SNPs within all of the regions covered by Lubbe *et al.* plus an additional 13 SNPS within eight regions). The current study of 45 SNPs within 38 regions includes SNPs within all of the regions covered by Hsu *et al.*, plus an additional 18 SNPs within 15 regions **(Supplementary Table 1)**. Using the methods described above, we estimated the proportion of the FRR that the panel of SNPs would explain for each of these previous studies. Given that each subsequent study includes SNPs within new regions as well as SNPs within the

same regions as previous studies, the estimated proportion of FRR explained by the SNPs in each subsequent study also increases.

Because the average FRR attributable to each SNP is not decreasing with increases in the number of SNPs **(Supplementary Table 1)**, there is evidence that the more recently discovered SNPs are as important as those previously identified, not less important. This suggests that, at least in the near future, there is potential for substantial increases in the predictive value of newly identified colorectal cancer SNPs for risk profiling. The total number of SNPs that may be eventually discovered is unknown. The SNPs identified to date have had to pass very stringent measures of association (very small p-values) to increase specificity, but this has severely limited the sensitivity
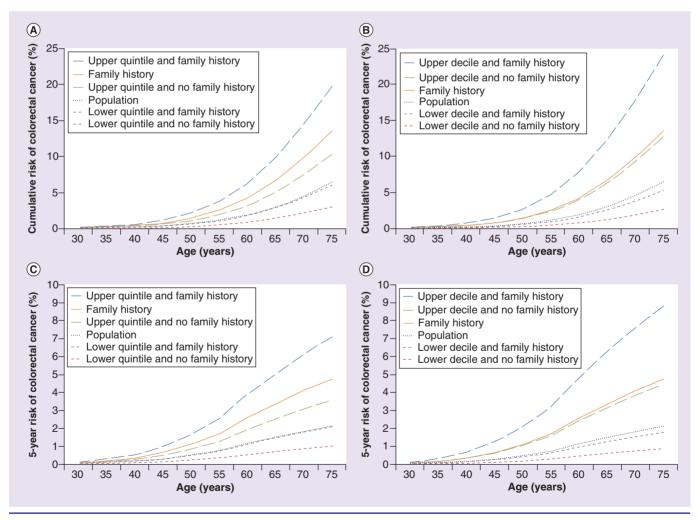

Figure 2. Australian risks of colorectal cancer (males and females combined) by age category, family history of colorectal cancer (first-degree relative) and by number of risk alleles. (A) Cumulative risks to age 70 years with highest and lowest quintiles for number of risk alleles. **(B)** Cumulative risks to age 70 years with highest and lowest deciles for number of risk alleles. **(C)** 5-year risks with highest and lowest quintiles for number of risk alleles. **(D)** 5-year risks with highest and lowest deciles for number of risk alleles.

of SNP detection studies. Relaxing this criteria or development of new methods to identify SNPs consistently associated with colorectal cancer risk might help uncover many more. One estimate of the total number of SNPs for colorectal cancer is 172 based on the observed associations and frequency of ten known SNPs [2]; but if the allele frequencies or strengths of association per risk allele for the as yet unknown SNPs are lower than for the current SNPs, a greater number of colorectal cancer SNPs would exist.

Using a different measure of SNP association on familial aggregation of the colorectal cancer, Jiao *et al.* [41] estimated that the combined heritability due to 31 SNPs (which were a subset of the SNPs assessed in the current study) was 0.65%. They estimated that if all SNPs with a minor allele frequency of at least 1% were to be used (over 200,000 SNPs), even if the association with colorectal cancer risk for these SNPs did not reach statistical significance, the heritability would reach 7.4%. In other words, the 31 SNPs accounted for only 9% of the total heritability of colorectal cancer that could be attributed to common variants suggesting there are additional SNPs to be identified.

A limitation of this study is the assumption that the associations of the risk alleles of colorectal cancer were assumed to be independent. This may not be true, for alleles of SNPs in close proximity, or even for alleles of SNPs on different chromosomes. Reasons for nonindependence might include linkage disequilibrium (although we have attempted to use only SNPs that previous studies have concluded as having independent effects) and SNPs being associated with causative genetic factors within the same biological pathway. If any of the risk alleles in this study were interacting, the actual risk of colorectal cancer may be higher or

lower than we have predicted. Another limitation is the lack of published studies assessing whether the strengths of associations for each SNP with colorectal cancer varies by age. Therefore, in absence of any such data, we have assumed that the associations were consistent with age.

Colorectal cancer screening programs advocate administering tests to individuals across apparently healthy populations to identify individuals who have either premalignant or early stages of colorectal cancer so that they may benefit from prevention or early treatment. In the average risk population, screening based on fecal occult blood testing reduces colorectal mortality by 15–25% [42]. Endoscopic screening can reduce mortality by 30–40% [43]. Ideally, deciding who should receive screening as well as the procedure and intensity of that screening should be based on the individual's risk of colorectal cancer. However, because there are currently no precise or valid methods to determine individual risk of the disease, targeted screening is only based on the very broad risk factors of age, gender, and sometimes, family history. This makes screening programs inefficient because many of those screened will never get colorectal cancer, and many of those not screened are at substantial risk of the disease [44]. Screening effectiveness and efficiency for all cancers could benefit from accurate and precise estimates of personal cancer risk, with specific screening recommendations relevant to that risk.

Conclusion

This study has provided evidence that the existing SNPs are useful for colorectal cancer risk profiling alone and in combination with family history of colorectal cancer. Based on the inability of the currently known risk-associated SNPs to

Table 2. Age at which the 5-year risk of colorectal cancer reaches or exceeds thresholds of 1%, for various categories of family history of colorectal cancer (at least one first-degree relative) and risk alleles of 45 single nucleotide polymorphisms.

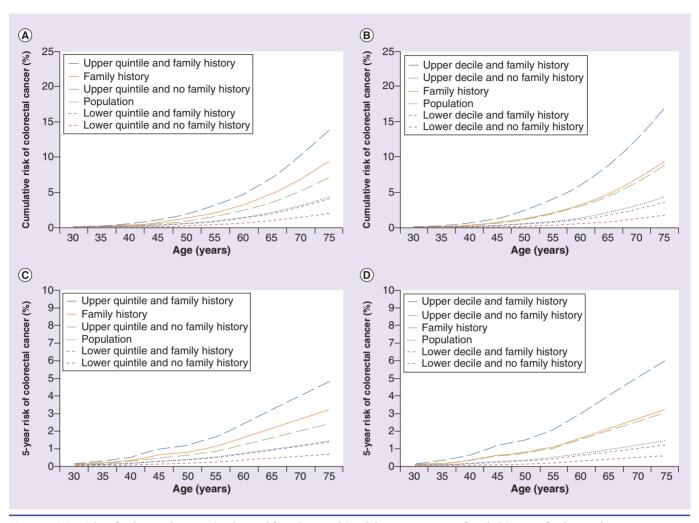| Risk category	USA (years)			Australia (years)		
	All	*Male*	*Female*	*All*	*Male*	*Female*
General population	70	67	73	63	61	71
Family history (first-degree relative)	58	55	61	53	52	59
Highest quintile of risk alleles	61	57	62	56	55	62
Highest decile of risk alleles	58	53	59	53	52	59
Family history and highest quintile	50	48	52	49	48	55
Family history and highest decile	48	46	48	47	46	53
Family history and lowest quintile	71	66	73	63	61	72
Family history and lowest decile	74	73	80	65	63	76

Figure 3. USA risks of colorectal cancer (males and females combined) by age category, family history of colorectal cancer (first-degree relative) and by number of risk alleles. (A) Cumulative risks to age 70 years with highest and lowest quintiles for number of risk alleles. **(B)** Cumulative risks to age 70 with highest and lowest deciles for number of risk alleles. **(C)** 5-year risks with highest and lowest quintiles for number of risk alleles. **(D)** 5-year risks with highest and lowest deciles for number of risk alleles.

explain more than a small proportion of familial aggregation in disease, more susceptibility SNPs are likely to exist and when identified, will lead to further improvements in colorectal cancer risk prediction. These findings, and their potential to improve the capacity of colorectal screening to reduce the burden of disease, point to the need for new statistical methods to examine the role for the vast number of SNPs that are associated with colorectal cancer but fail to meet current thresholds of significance.

Future perspective

Larger, and therefore more statistically powerful GWAS will identify additional SNPs not included in this analysis. In addition, more sophisticated statistical methods are being

developed that will identify additional risk-associated SNPs for colorectal cancer. Adding these SNPs, when discovered, to risk profile models such as ours, will increase the ability to identify those who would benefit most from colorectal cancer screening. At the same time, the costs for tests are decreasing and new technology is being developed that will soon enable individual, on demand testing for SNPs. This opens up the possibility that within 5–10 years primary care physicians will be able to test and advise screening practices within a consultation.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine. com/doi/full/10.2217/fon.15.303

EXECUTIVE SUMMARY

What is already known about this subject?

- Colorectal cancer risk varies from person to person, and there are various screening modalities that can be recommended depending on individual risk.

- Inherited genetic risk factors are known for colorectal cancer.

- Some of these genetic factors can be measured by common genetic markers called single nucleotide polymorphisms (SNPs) that are weakly associated with colorectal cancer risk.

What are the new findings?

- In combination, the 45 known susceptibility SNPs can be used to define risk categories for colorectal cancer.

- These risk categories are at least as important in terms of differentiating risk as having a family history of the disease.

- A substantial proportion of the risk prediction from SNPs is independent of family history. This means SNPs and family history can be combined and used as predictors of risk.

How might it impact on clinical practice in the foreseeable future?

- If these results are verified, and testing for these genetic markers can be shown to be cost-effective, then testing for these genetic markers could become a routine test for evaluating screening recommendations especially for people with a family history.

References

Papers of special note have been highlighted as:
• of interest

1   Vasen HF, Van Der Meulen-De Jong AE, De Vos Tot Nederveen Cappel WH, Oliveira J, Group EGW. Familial colorectal cancer risk: ESMO clinical recommendations. *Ann. Oncol.* 20(Suppl. 4), 51–53 (2009).

2   Tenesa A, Dunlop MG. New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat. Rev. Genet.* 10(6), 353–358 (2009).

•   **Provided an estimation of the potential number single nucleotide polymorphisms (SNPs) to explain the remaining excess familial risk for colorectal cancer.**

3   Welter D, Macarthur J, Morales J *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 42, D1001–D1006 (2014).

4   Whiffin N, Dobbins SE, Hosking FJ *et al.* Deciphering the genetic architecture of low-penetrance susceptibility to colorectal cancer. *Hum. Mol. Genet.* 22(24), 5075–5082 (2013).

5   Zhang B, Jia W-H, Matsuda K *et al.* Large-scale genetic study in east Asians identifies six new loci associated with colorectal cancer risk. *Nat. Genet.* 46(6), 533–542 (2014).

6   Michailidou K, Hall P, Gonzalez-Neira A *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* 45(4), 353–361, 361e1–361e2 (2013).

7   Al Olama AA, Kote-Jarai Z, Berndt SI *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat. Genet.* 46(10), 1103–1109 (2014).

8   Dunlop MG, Tenesa A, Farrington SM *et al.* Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. *Gut* 62(6), 871–881 (2013).

•   **First study attempting to calculate risk score for colorectal cancer from measured SNP data.**

9   Purcell S, Neale B, Todd-Brown K *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81(3), 559–575 (2007).

10 PLINK.
    http://pngu.mgh.harvard.edu/purcell/plink/

11 Lubbe SJ, Di Bernardo MC, Broderick P, Chandler I, Houlston RS. Comprehensive evaluation of the impact of 14 genetic variants on colorectal cancer phenotype and risk. *Am. J. Epidemiol.* 175(1), 1–10 (2012).

12 Australian Institute of Health and Welfare (AIHW). Australian Cancer Incidence and Mortality (ACIM) books: Colorectal cancer (also called bowel cancer) Canberra: AIHW (2015).
    www.aihw.gov.au/acim-books/

13 Howlader N, Noone AM, Krapcho M *et al.* SEER Cancer Statistics Review, 1975–2012, National Cancer Institute, Bethesda, MD, USA.
    http://seer.cancer.gov/csr/1975_2012

14 Antoniou AC, Easton DF. Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet. Epidemiol.* 25(3), 190–202 (2003).

15 Win AK, Dowty JG, Cleary SP *et al.* Risk of colorectal cancer for carriers of mutations in *MUTYH*, with and without a family history of cancer. *Gastroenterology* 146(5), 1208–1211, e1201–e1205 (2014).

16 Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am. J. Gastroenterol.* 96(10), 2992–3003 (2001).

17 Zhang B, Jia WH, Matsuo K *et al.* Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in east Asians. *Int. J. Cancer* 135(4), 948–955 (2014).

18 Dunlop MG, Dobbins SE, Farrington SM *et al.* Common variation near *CDKN1A*, *POLD3* and *SHROOM2* influences colorectal cancer risk. *Nat. Genet.* 44(7), 770–776 (2012).

19 Spain SL, Carvajal-Carmona LG, Howarth KM *et al.* Refinement of the associations between risk of colorectal cancer and polymorphisms on chromosomes 1q41 and 12q13.13. *Hum. Mol. Genet.* 21(4), 934–946 (2012).

20 Haerian MS, Baum L, Haerian BS. Association of 8q24.21 loci with the risk of colorectal cancer: a systematic review and meta-analysis. *J. Gastroenterol. Hepatol.* 26(10), 1475–1484 (2011).

21 Tomlinson IPM, Carvajal-Carmona LG, Dobbins SE *et al.* Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet.* 7(6), e1002105 (2011).

22 Al-Tassan NA, Whiffin N, Hosking FJ *et al.* A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. *Sci. Rep.* 5, 10442 (2015).

23 Peters U, Hutter CM, Hsu L *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum. Genet.* 131(2), 217–234 (2012).

24 Whiffin N, Hosking FJ, Farrington SM *et al.* Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum. Mol. Genet.* 23(17), 4729–4737 (2014).

25 Houlston RS, Cheadle J, Dobbins SE *et al.* Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.* 42(11), 973–977 (2010).

26 Peters U, Jiao S, Schumacher FR *et al.* Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology* 144(4), 799–807 e724 (2013).

27 Schumacher F, Stenzel S, Jiao S *et al.* Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat. Commun.* 6, 7138 (2015).

28 Real LM, Ruiz A, Gayan J *et al.* A colorectal cancer susceptibility new variant at 4q26 in the Spanish population identified by genome-wide association analysis. *PLoS ONE* 9(6), e101178 (2014).

29 Schmit SL, Schumacher FR, Edlund CK *et al.* A novel colorectal cancer risk locus at 4q32.2 identified from an international genome-wide association study. *Carcinogenesis* 35(11), 2512–2519 (2014).

30 Jia WH, Zhang B, Matsuo K *et al.* Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat. Genet.* 45(2), 191–196 (2013).

31 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.* 40(5), 623–630 (2008).

32 Zanke BW, Greenwood CM, Rangrej J *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* 39(8), 989–994 (2007).

33 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.* 39(8), 984–988 (2007).

34 Kocarnik JD, Hutter CM, Slattery ML *et al.* Characterization of 9p24 risk locus and colorectal adenoma and cancer: gene-environment interaction and meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* 19(12), 3131–3139 (2010).

35 Tenesa A, Campbell H, Theodoratou E *et al.* Common genetic variants at the MC4R locus are associated with obesity, but not with dietary energy intake or colorectal cancer in the Scottish population. *Int. J. Obes. (Lond.)* 33(2), 284–288 (2009).

36 Tomlinson IP, Carvajal-Carmona LG, Dobbins SE *et al.* Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet.* 7(6), e1002105 (2011).

37 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.* 40(12), 1426–1435 (2008).

38 Ryan BM, Wolff RK, Valeri N *et al.* An analysis of genetic factors related to risk of inflammatory bowel disease and colon cancer. *Cancer Epidemiol.* 38(5), 583–590 (2014).

39 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.* 39(11), 1315–1317 (2007).

40 Hsu L, Jeon J, Brenner H *et al.* A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology* 148(7), 1330.e14–1339.e14 (2015).

•   **Most recent and largest study to calculate risk score for colorectal cancer from measured SNP data.**

41 Jiao S, Peters U, Berndt S *et al.* Estimating the heritability of colorectal cancer. *Hum. Mol. Genet.* 23(14), 3898–3905 (2014).

42 Hewitson P, Glasziou P, Irwig L, Towler B, Watson E. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database Syst. Rev.* (1), CD001216 (2007).

43 Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ* 348, g2467 (2014).

44 Ait Ouakrim D, Boussioutas A, Lockett T *et al.* Screening practices of unaffected people at familial risk of colorectal cancer. *Cancer Prev. Res. (Phila.)* 5(2), 240–247 (2012).