Germline susceptibility to colorectal cancer due to base-excision repair gene defects

Am. J. Hum. Genet. 77:112–119, 2005

Germline Susceptibility to Colorectal Cancer Due to Base-Excision Repair Gene Defects

Susan M. Farrington,[1,4,*] Albert Tenesa,[1,4,*] Rebecca Barnetson,[1,4] Alice Wiltshire,[1,4] James Prendergast,[1,4] Mary Porteous,[1,2] Harry Campbell,[1,3] and Malcolm G. Dunlop[1,2]

[1]Colon Cancer Genetics Group, School of Clinical and Molecular Medicine, [2]Clinical Genetics Department, and [3]Public Health Sciences, University of Edinburgh, and [4]Medical Research Council (MRC), Human Genetics Unit, Edinburgh

DNA repair is a key process in the maintenance of genome integrity. Here, we present a large, systematically collected population-based association study (2,239 cases; 1,845 controls) that explores the contribution to colorectal cancer incidence of inherited defects in base-excision repair (BER) genes. We show that biallelic *MUTYH* defects impart a 93-fold (95% CI 42–213) excess risk of colorectal cancer, which accounts for 0.8% of cases aged <55 years and 0.54% of the entire cohort. Penetrance for homozygous carriers was almost complete by age 60 years. Significantly more biallelic carriers had coexisting adenomatous polyps. However, notably, 36% of biallelic carriers had no polyps. Three patients with heterozygous *MUTYH* defects carried monoallelic mutations in other BER genes (*OGG1* and *MTH1*). Recessive inheritance accounted for the elevated risk for those aged <55 years. However, there was also a 1.68-fold (95% CI 1.07–2.95) excess risk for heterozygous carriers aged >55 years, with a population attributable risk in this age group of 0.93% (95% CI 0%–2.0%). These data provide the strongest evidence to date for a causative role of BER defects in colorectal cancer etiology and show, to our knowledge for the first time, that heterozygous *MUTYH* mutations predispose to colorectal cancer later in life. These findings have clinical relevance for BER gene testing for patients with colorectal cancer and for genetic counseling of their relatives.

Introduction

The role of base-excision repair (BER) in the maintenance of genome stability is primarily to counter oxidative DNA damage, which generates 8-oxoguanine products (8-oxoG). In humans, MYH (MIM 604933), OGG1 (MIM 601982), and MTH1 (MIM 600312) function in consort to identify and repair 8-oxoG incorporated into DNA, as well as to remove modified nucleoside. Recent studies have identified biallelic germline defects in *MUTYH,* in a proportion of families with multiple colorectal polyposis, that are not due to germline *APC* (MIM 175100 and 608456) mutations (Al-Tassan et al. 2002; Sieber et al. 2003). Although these studies provide indirect evidence, it is important to establish whether BER gene defects predispose to colorectal cancer (MIM 114500), to estimate the level of associated risk, and to determine the attributable contribution of such defects to overall disease incidence. Previous studies have provided some supporting evidence that biallelic mutations are associated with excess cancer risk (Croitoru et al. 2004; Fleischmann et al. 2004). Here, we present an analysis of the largest cohort study to date, thereby affording the opportunity to assess the effect of homozygous and heterozygous BER gene mutations on colorectal cancer risk. We assembled a large, systematically recruited prospective cohort of patients from across Scotland, shortly after diagnosis of colorectal cancer and irrespective of family history. We also systematically recruited healthy Scottish population control individuals through the central National Health Service (NHS). Using this population-based resource, we set out to determine, by a genetic association strategy, the role of BER genes in colorectal cancer susceptibility.

Received January 18, 2005; accepted for publication April 14, 2005; electronically published May 3, 2005.

Address for correspondence and reprints: Dr. Susan M, Farrington, Colon Cancer Genetics Group, MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, United Kingdom. E-mail: Susan.Farrington@hgu.mrc.ac.uk

* These two authors contributed equally to this work.

0002-9297/2005/7701-0011$15.00

Subjects and Methods

Assembly of the Cohort and Sample Collection

A populationwide accrual of prospective colorectal cancer cases has been in progress since 1999. Cases are ascertained through direct contact with every surgical and pathology department in Scotland. All cases had histologically confirmed adenocarcinoma of the colon or rectum. Blood DNA samples were obtained from patients after counseling and receipt of informed consent. Population-based and age- and sex-matched controls were

systematically identified, and blood DNA samples were obtained from them. A questionnaire about family information and lifestyle and medical history was completed for patients and controls. Dietary risk-exposure data were also collected by use of a validated food-frequency questionnaire. For cases, tumor stage, pathology, clinical presentation, as well as the presence of synchronous polyps were documented. These studies are subject to all necessary approvals from local ethics research committee (LREC), multicenter research ethics committee (MREC), and NHS research and development management in every participating hospital.

Analysis of MUTYH *Variants Y165C and G382D*

A two-stage approach was used to efficiently identify subjects carrying heterozygous or homozygous BER gene variants. In the first step, assays were designed using Primer Express v2.0 software (Applied Biosystems [AB]) for the two commonly reported variants, Y165C and G382D, in patients with multiple polyposis. Allelic discrimination for each variant employed allele-specific *Taq*Man MGB probes (AB), resolved on an ABI 7900 Analyzer, by use of SDS v2.1 software. Probe and primer details are available on request. Each Y165C and G382D variant identified by the *Taq*Man approach was confirmed by repeat DNA sequence analysis. In the second phase, all subjects heterozygous for Y165C and G382D underwent sequence analysis of the entire coding region of *MUTYH,* and the heterozygote cases were also screened for *OGG1* and *MTH1* gene variants. GenBank accession numbers for the genes are AF527839, NT_022517, and NT_007819, respectively. PCR products were treated with Exonuclease 1, shrimp alkaline phosphatase (Amersham Biosciences) and then were sequenced using ABI Big Dye terminator V3.0 chemistry, with precipitated products separated on an ABI 3700. Details of primer sequences are available on request. Using this approach, we identified all homozygotes or compound heterozygotes and any subjects in whom at least one allele was either Y165C or G382D. However, it should be noted that we could have missed some biallelic carriers who did not have these two common variants. Hence, if anything, our estimates may underrepresent the contribution of BER genes, but we consider this to be a marginal effect.

Assessment of Variants of BER Genes

Each variant identified by sequencing was confirmed by repeat sequence analysis. Allele frequency for each newly identified variant was then determined in at least 340 control chromosomes, to confirm that the variant was not simply a common polymorphism. Any common polymorphic variants were discarded, after which each identified variant was subjected to rigorous bioinformatic analysis. We used SIFT, which predicts deleterious coding variants on the basis of cross-species conservation; PolyPhen, which predicts the effect by use of conservation and any protein structure available in the public domain; T-Coffee, which aligns closely related Ensembl orthologues; protein domains predicted using Pfam; assessment of potential splicing effects by use of GENSCAN; ESEfinder; and Berkeley *Drosophila* Genome Project splice-site prediction.

Functional Analysis of MUTYH *nt 9639 a→g Variant*

A variant was identified that was predicted to affect splicing (*MUTYH*; nt 9639 a→g). In this case, RNA was extracted from blood leukocytes by use of Tri-reagent (Sigma) and was processed to cDNA (Boeringer Mannheim First Strand Synthesis kit). A transcript product of 475 bp was amplified using exonic primers in exons 10 and 14 (primer details available on request). The product was cloned using the Topo cloning kit (Invitrogen), and positive colonies were amplified and sequenced. Similarly, a genomic amplicon covering the nt 9639 a→g change and the G382D locus was cloned to demonstrate recessive inheritance of the two variants.

Test for Association

Association was tested using Fisher's exact test in a 3 × 2 table of genotype by colorectal cancer status, thereby making no prejudgment on potential mode of inheritance.

Estimate of Genotype Relative Risk

As an adaptation of the method of Hugot et al. (2001), the genotype relative risk (GRR) is defined as follows:

$$\mathrm{GRR}(AA) = \frac{\Pr(D|AA)}{\Pr(D|AA)} = 1 ,$$

$$\mathrm{GRR}(Aa) = \frac{\Pr(D|Aa)}{\Pr(D|AA)} = \frac{\Pr(Aa|D) \times \Pr(AA)}{\Pr(AA|D) \times \Pr(Aa)} ,$$

and

$$\mathrm{GRR}(aa) = \frac{\Pr(D|aa)}{\Pr(D|AA)} = \frac{\Pr(aa|D) \times \Pr(AA)}{\Pr(AA|D) \times \Pr(aa)} ,$$

where $\Pr(G|D)$ is the frequency of genotype G among cases and $\Pr(G)$ is the expected Hardy-Weinberg equilibrium proportion, as obtained from the allele frequency in the control population. The control population was assumed to be a representative sample of the general population. The 95% CIs for the GRRs were obtained

Table 1

Population GRR Associated with MYH Y165C

Population and Genotype	No. of Subjects with Genotype		GRR	95% CI
	Case	Control		
Entire cohort (P = .228):				
GG	0	0	.00	.00–.00
AG	17	8	1.76	.90–4.16
AA	2,202	1,824	1.00	1.00–1.00
U55 (P = 1.000):				
GG	0	0	.00	.00–.00
AG	7	4	1.08	.37–3.72
AA	867	535	1.00	1.00–1.00
O55 (P = .180):				
GG	0	0	.00	.00–.00
AG	10	4	2.42	.96–8.71
AA	1,335	1,289	1.00	1.00–1.00

by bootstrapping 1,000 independent samples. For each of the samples, the GRR was estimated, then the same 1,000 samples were ordered within each genotype, and the 50th and 950th estimates were taken as the lower and upper limits of the 95% CI.

Penetrance Estimates

Penetrance for colorectal cancer at a given age is defined as the probability that a randomly selected individual with genotype G will develop the disease by that age, with the assumption that that individual does not die of another cause. If A_x is the event "affected at age x:$x + 1$, given being disease-free at x," then the probability that an individual is affected by age x, given the genotype G, is $\Pr(A_x|G)$. Similarly if $\overline{A}_x$ is the event "not affected at age x:$x + 1$, given being disease-free at x," then the probability that an individual is not affected by age x, given genotype G, is $\Pr(\overline{A}_x|G)$. Population incidence rates, $\Pr(A_x)$, were obtained from Scottish Health Statistics, and the sex-average incidence rates were used.

Then

$$\frac{\Pr(A_x|G)}{\Pr(\overline{A}_x|G)} = \frac{\Pr(G|A_x) \times \Pr(A_x)}{\Pr(G|\overline{A}_x) \times \Pr(\overline{A}_x)} ,$$

$$\frac{\Pr(A_x|G)}{1 - \Pr(A_x|G)} = \frac{\Pr(G|A_x) \times \Pr(A_x)}{\Pr(G|\overline{A}_x) \times [1 - \Pr(A_x)]} ,$$

and

$$\Pr(A_x|G) = \frac{\frac{\Pr(G|A_x) \times \Pr(A_x)}{\Pr(G|\overline{A}_x) \times [1-\Pr(A_x)]}}{1 + \frac{\Pr(G|A_x) \times \Pr(A_x)}{\Pr(G|\overline{A}_x) \times [1-\Pr(A_x)]}} .$$

There were not enough observations for each age, so $\Pr(G|A_x)$ was estimated using logistic regression (i.e., age was used as a predictor of genotype). Similarly, $\Pr(G|\overline{A}_x)$ was estimated from controls. The penetrance for genotype G is then defined as:

$$P_G = 1 - \prod_1^x [1 - \Pr(A_x|G)] .$$

Estimate of CIs

The 95% CI for the estimate of the penetrance was obtained by bootstrapping (i.e., by sampling with replacement) samples of cases and controls. A total of 500 independent samples were obtained of size equal to the total number of cases and controls. Penetrance was estimated for each sample, and mean penetrance over rep-

Table 2

Population GRR Associated with MYH G382D

Population and Genotype	No. of Subjects with Genotype		GRR	95% CI
	Case	Control		
Entire cohort (P = .010):				
AA	8	0	121.23	44.48–325.10
GA	35	20	1.46	.93–2.43
GG	2,181	1,808	1.00	1.00–1.00
U55 (P = .402):				
AA	4	0	146.69	26.71–998.71
GA	12	6	1.24	.54–3.42
GG	864	531	1.00	1.00–1.00
O55 (P = .049):				
AA	4	0	102.19	22.09–333.10
GA	23	14	1.60	.93–2.95
GG	1,317	1,277	1.00	1.00–1.00

Table 3

Population GRR Associated with *MUTYH* Gene

Population and Genotype[a]	No. of Subjects with Genotype		GRR	95% CI
	Case	Control		
Entire cohort (P = .001):				
MM	12	0	92.65	41.60–213.20
WM	45	28	1.35	.92–2.07
WW	2,160	1794	1.00	1.00–1.00
U55 (P = .976):				
MM	7	0	91.73	22.53–293.41
WM	14	10	.87	.43–1.72
WW	851	523	1.00	1.00–1.00
O55 (P = .014):				
MM	5	0	77.26	20.51–208.98
WM	31	18	1.68	1.07–2.95
WW	1,309	1,271	1.00	1.00–1.00

[a] W = wild-type allele; M = mutant allele.

licates and standard deviation (SD) were obtained. The 95% CI was calculated using the mean ± 1.96 SD.

Results

Analysis of MUTYH *Variants Y165C and G382D*

All variants detected by *Taq*Man analysis were confirmed by genomic sequencing. Eight (0.36%) of the patients were homozygous for G382D. There were no homozygotes for Y165C, but three Y165C/G382D compound heterozygotes were identified. There were no biallelic defects in any control samples. Association was first tested, using Fisher's exact test in a 3 × 2 table of genotype by colorectal cancer status, thereby making no prejudgment on potential mode of inheritance. Next, GRR was calculated using the method of Hugot et al. (2001), because the usual method, estimated by odds ratio, is not possible because there were no homozygote controls. The frequency of Y165C and G382D alleles in cases and controls is presented in tables 1 and 2, respectively.

The data presented in table 2 establish the fact that the G382D locus is associated with colorectal cancer in the complete cohort (P = .0104; GRR 121; 95% CI 44–325 for the G382D/G382D genotype). To explore any age-specific effect, the cohort aged <55 years at diagnosis (U55) was compared with those aged >55 years at diagnosis (O55) (table 2). Because of the lower number of U55 subjects, there was no statistically significant association with G382D for the U55 cohort, whereas association for the O55 cohort remained statistically significant (P = .0494). There was no association with the Y165C locus (table 1), possibly because of the rarity of the mutant variant in the Scottish population. We consider the lack of association in the early-onset cohort to be due primarily to the low allele frequency of these individual variants and the resultant lack of statistical power, even though this study involved large case and control cohorts.

Significance of Other Variants in BER Genes

In light of prior genetic and functional evidence of an additive effect of G382D and Y165C alleles (Al-Tassan et al. 2002; Sieber et al. 2003), we next determined the overall prevalence of biallelic defects, using a pragmatic approach. To find second BER gene defects, DNA from all carriers of monoallelic Y165C or G382D mutations (n = 74) was sequenced for each exon and intron/exon boundary of *MUTYH*. BER genes *OGG1* and *MTH1* were also sequenced in cases with heterozygous Y165C or G382D MYH alleles. All identified variants were excluded as polymorphisms by confirmation of wild-type sequence in at least 340 control chromosomes and likely functional relevance assessed by bioinformatic analysis, as described in the "Subjects and Methods" section.

In addition to the eight G382D/G382D homozygotes and the three Y165C/G382D compound heterozygotes described above, we identified a further patient with a heterozygous G382D mutation who had a second genomic *MUTYH* defect (nt 9639 a→g) that we consider to be pathogenic. This variant was confirmed to reside on the wild-type allele by genomic DNA cloning and sequencing. The variant was predicted to affect splicing. This was confirmed by cDNA analysis, which showed only mutant G382D transcript and no wild-type transcript. Thus, the patient was hemizygous for G382D at the RNA level. The *Taq*Man approach used for the G382D variant identified all nt 9639 a→g variants, so we have systematically screened all samples for this variant. Thus, the genotyping strategy would have identified the *MUTYH* defect nt 9639 a→g variant with equal sensitivity in cases and controls. Hence, we have included

this variant in further analyses of the effect of overall *MUTYH* genotype, for a total of 12 biallelic carriers (all cancer patients) and 73 subjects (45 patients and 28 controls) with monoallelic *MUTYH* alleles that we are confident are pathogenic (table 3).

We next considered other samples that might have a second BER gene defect. In all, there were an additional three samples with heterozygous G382D or Y165C alleles that had another BER gene defect that we consider likely to be pathogenic, but we cannot confirm or refute this. Hence, we did not include these other BER alleles in the analysis of the overall effect of putative BER gene defects (table 3). Two patients with Y165C or G382D mutations (one of each) had a second allelic variant in MTH1 (R31Q) that has been reported elsewhere in multiple polyposis (Sieber et al. 2003). Codon 31 of MTH1 is evolutionarily conserved, and bioinformatics interrogation suggests that the R31Q mutation affects protein function. Furthermore, we identified the variant in a total of 2 of 84 patient chromosomes, compared with 0 of 368 control chromosomes, which suggests that this is not a common polymorphic variant in the population (although we cannot directly compare R31Q prevalence, because these are two different groups). Taken together with previously published studies, we consider that the MTH1 R31Q variant is likely to be functionally important and consequently is likely to be a pathogenic mutation. One patient with an MTH1 R31Q variant also had a P391L variant in MYH that is at a highly conserved residue and is predicted to affect protein function. However, the significance of this variant is unclear, in light of the above discussion. A third patient with a monoallelic MYH G382D allele also had a variant in OGG1 (R197W). The variant is at a highly conserved codon, and bioinformatics analysis predicts a profound effect on protein function; again, the significance of this cannot currently be determined.

Analysis of the MUTYH *Gene and Colorectal Cancer Risk*

To assess biallelic inactivation of the entire gene, we calculated the combined risk for all significant *MUTYH* variants that had been analyzed in the complete cohort. This analysis included eight G382D/G382D homozygotes, three Y165C/G382D compound heterozygotes, and one nt 9639 a→g /G382D compound heterozygote, as discussed above. The data presented in table 3 convincingly establish the fact that the *MUTYH* gene is significantly associated with colorectal cancer ($P = .0012$). Biallelic inactivation imparts an overall 93-fold excess risk (GRR 93; 95% CI 42–213). As was the case for analysis of the G382D variant alone, the reduction in numbers for the U55 cohort resulted in loss of statistical power to detect association with the gene in that group, although the O55 group remains significant ($P = .0138$).

Separation of the cohorts by age reveals an age-specific risk effect and shows, for the first time, a significant monoallelic effect for late-onset disease when the empirical CIs produced by bootstrap analysis are used. There was a 1.68-fold excess risk (95% CI 1.07%–2.95) for heterozygous carriers aged >55 years and a population-attributable risk in this age group of 0.93% (95% CI 0%–2.0%). However, these results should be taken with some caution, because the significance was marginal, despite the fact that we have studied large case/control cohorts. Furthermore, analysis by use of standard CIs for the odd ratios in testing for association gave an almost significant result ($P = .085$; 95% CI 0.93–3) for a monoallelic effect. It was only with bootstrap analysis that this effect was significant at the 5% level. Nonetheless, these are novel observations that suggest that heterozygous *MUTYH* variants impart a modest increased risk later in life.

Estimate of Penetrance for MUTYH *Gene Defects*

We first estimated penetrance for G382D alleles at the *MUTYH* locus using the method of Satagopan et al. (2001), modified for application to a recessive trait. Figure 1 shows age-specific penetrance and 95% CIs for the G382D/G382D genotype alone. Of G382D/G382D carriers, ~55% developed colorectal cancer by age 40 years, and cancer had developed in all G382D/G382D carriers by age 65 years. Next, we estimated age-specific penetrance for all systematically determined biallelic defects identified in *MUTYH* (fig. 2). This analysis suggests that biallelic inactivation of the gene is highly penetrant, with all homozygous carriers developing cancer by age 60 years.

These findings establish the fact that there is a substantially elevated colorectal cancer risk early in life for people with biallelic MYH defects.

Synchronous Polyps in Biallelic MUTYH *Cancer Patients*

Data on polyp prevalence in cancer patients by genotype are presented in table 4. It is noteworthy that 4 (36%) of the 11 biallelic carriers with cancer for whom we had reliable polyp information had no concurrent metaplastic or adenomatous polyps. This emphasizes the fact that using polyp presence as a surrogate to enrich for *MUTYH* mutation carriers is quite insensitive. The data presented in table 4 show that there was a clear relationship between synchronous polyps and *MUTYH* genotype (all polyp types $P = .025$; adenomatous polyps $P = .007$; multiple adenomas $P < .0001$ [with use of Fisher's Exact Test]). In all, 64% of biallelic carriers had polyps; in all cases, these were multiple adenomas.

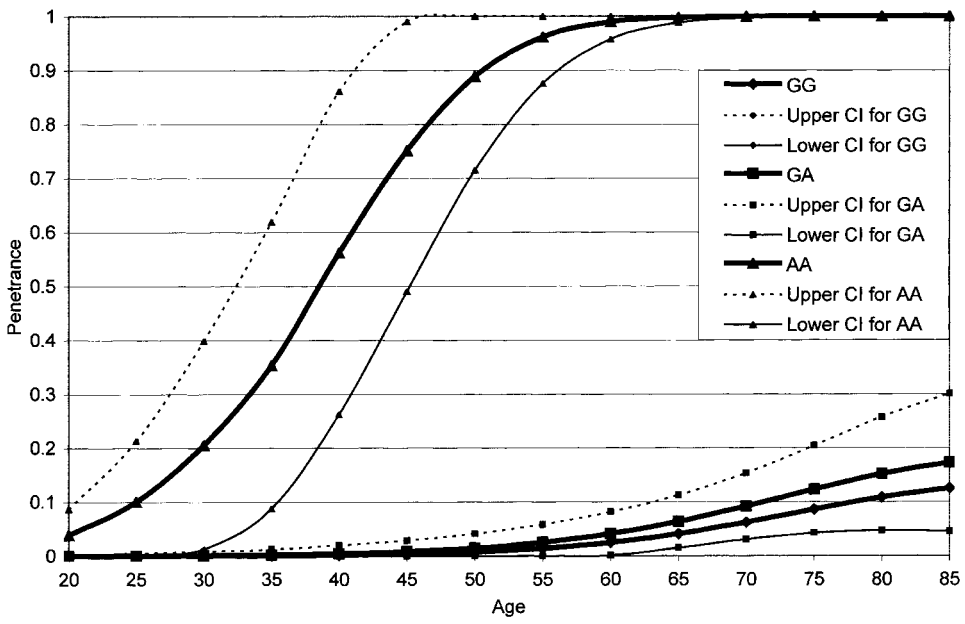

Figure 1 Penetrance curve of the MYH G382D locus. Age-related penetrance was estimated and plotted for all three genotypes, including 95% CIs.

The same analyses were performed, but with comparison of only the heterozygotes and wild-type genotypes, using 2 × 2 contingency tables; there were no significant results for presence of polyps with a heterozygote genotype ($P > .24$).

Discussion

Taken together, these data establish conclusively that defects in BER genes predispose to colorectal cancer. The findings provide robust evidence of causal involvement of biallelic BER gene defects in early-onset colorectal cancer, given the fact that there was complete penetrance by age 60 years. These findings substantially extend previous indirect evidence of involvement in colorectal cancer susceptibility through study of families with multiple polyposis (Al-Tassan et al. 2002; Sieber et al. 2003) and, combined with our findings of an excess of adenomatous polyps, suggest that polyps in such cases are premalignant. Our findings are also supported by previous suggestive evidence from smaller studies that lacked power to definitively establish whether there is an association between biallelic *MUTYH* variants and colorectal cancer (Enholm et al. 2003; Fleischmann et al. 2004; Wang et al. 2004).

There is only one previous study (Croitoru et al. 2004) that has shown an association between biallelic *MUTYH* variants and colorectal cancer. In that study, the authors proposed indirect evidence of a heterozygous effect because they observed nonrandom loss of heterozygosity of the wild-type allele in colorectal cancer. The findings presented here do, in fact, support their view and provide the first evidence of a modest colorectal cancer risk associated with heterozygous *MUTYH* mutations later in life. It is noteworthy that we found evidence of a significant excess risk associated with heterozygous *MUTYH* gene defects only when we considered the entire gene effect, and, importantly, this effect emerged only for late-onset disease. The effect was small and only just significant, and the results should be taken with some caution. Bootstrapping CIs are generally more robust than asymptotic CIs; however, it must be noted that asymptotic CIs for the odds ratio gave a slightly less significant result for the same data set. It is also possible that some of the excess risk we are detecting in heterozygotes is due to variants on the other allele that we have not detected. Data from mouse models indicates that, on an $Apc^{Min/+}$ background, monoallelic *Myh* inactivation does not increase tumor burden or the signature G-T transversions of the remaining *Apc* allele in the mouse tumors (Sieber et al. 2004). However, the study presented here is the largest to date, and previous studies have concentrated on early-onset disease or mixed cohorts (Fleischmann et al. 2004; Wang et al. 2004). Thus, only studies involving even larger numbers of later-onset cancer cases might be able to confirm our evidence of a monoallelic effect.

The study presented here emphasizes the requirement for very large population-based cohorts for robust assessment of the role of putative colorectal cancer sus-

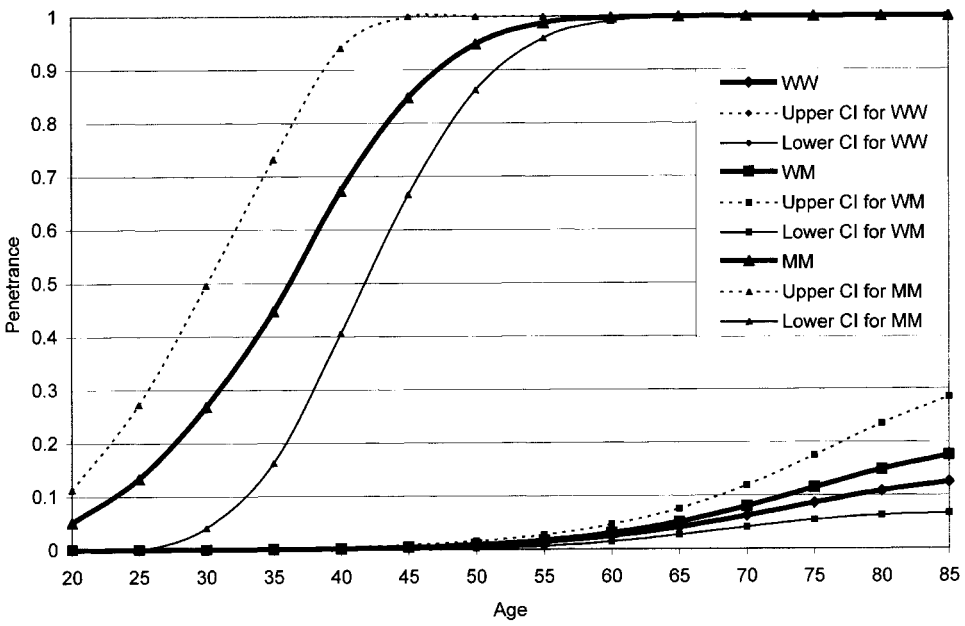

Figure 2 Penetrance curve of the *MUTYH* gene. All systematically verified and functionally deficient variants were included in the calculations, to estimate the penetrance of the entire gene. This therefore includes variants G382D, Y165C, and nt 9639 a→g. W = wild-type allele; M = mutant allele.

ceptibility alleles by genetic association strategies, even when relative risk is high. The relatively low allele frequency is likely to be a feature of many other cancer susceptibility alleles and raises some concerns for future power to detect associations where there is no prior biological hypothesis behind locus selection.

For clinical purposes, it is interesting to note that more than one-third of biallelic carriers did not have any synchronous polyps. This emphasizes the fact that use of polyps as an approach to enrich for *MUTYH* carriers with cancer is quite insensitive, in contrast with the findings of *APC*-negative multiple polyposis cases (Sampson et al. 2003; Sieber et al. 2003). Overall, there was an excess of synchronous polyps in biallelic carriers compared with others in this cohort, especially when adenomatous lesions were considered. Interestingly, monoallelic carriers did not appear to have an excess of polyps, suggesting the possibility that the excess cancer risk for heterozygotes later in life might be through a different mechanism.

Data presented here indicate that 1 of 50 patients who are diagnosed with colorectal cancer at age <40 years and 1 of 150 patients aged <55 years carry biallelic mutations in BER genes that are causally linked to cancer development. In defining the mutation carrier frequency for *MUTYH* and other BER genes, the present work informs future decisions about offering genetic testing to patients with early-onset colorectal cancer. Furthermore, the findings also have substantial clinical importance for the siblings of carriers, who have at least a 1/4 risk of colorectal cancer by age <60 years, by nature of the recessive genetic trait segregating in the family. Very large studies of late-onset disease-carrier status are required to replicate or refute the evidence

Table 4

***MUTYH* Genotype and the Presence of Synchronous Benign Polyps**

	No. (%) of Subjects with				
Genotype	No Polyps	All Polyp Types	Multiple Polyps	Adenomas	Multiple Adenomas
MM (n = 11)	4 (36)	7 (64)	7 (64)	7 (64)	7 (64)
WM (n = 44)	31 (70)	13 (30)	5 (11)	8 (18)	3 (7)
WW (n = 1,167)	859 (74)	308 (26)	100 (9)	224 (19)	54 (5)

Note.—Adenomas were either tubular or tubulovillous subtypes. Other polyps were metaplastic/hyperplastic. Three or more polyps were categorized as "multiple." Reliable concurrent polyp-prevalence data were not available for all patients with cancer, so the numbers for whom information was available are provided for each genotype.

presented here that monoallelic *MUTYH* defects contribute to colorectal cancer incidence.

Acknowledgments

We gratefully acknowledge the participation of all patients and control individuals, without whom this work would not have been possible. We thank all of the nursing and office staff employed by the Colorectal Cancer Genetic Susceptibility Study and the MRC Scottish Colorectal Cancer Study, for their tireless work in recruitment, which has been a major logistic endeavor. We also acknowledge the collaborative environment that we have enjoyed with NHS consultant surgeons and nursing teams in every Scottish hospital. We also thank the relevant departments in central Scottish NHS, including Cancer Registry, the Scottish Cancer Intelligence Unit of the Information and Statistics Division, and the Family Practitioner Committee, for invaluable help in recruiting population controls. We also thank Andrew Carothers for statistical advice. Current ethical approvals from MREC and LREC for all aspects of the study are held by M.P., H.C., and M.G.D. The work is funded by Chief Scientist Office grant CZB/4/94, Cancer Research UK grant C348/A3758, and MRC grant G0000657-53203. J.P. is funded by MRC Bioinformatics Research Studentship G74/93.

Web Resources

Accession numbers and URLs for data presented herein are as follows:

Berkeley *Drosophila* Genome Project, http://www.fruitfly.org/seq_tools/splice.html
ESEfinder, http://rulai.cshl.edu/tools/ESE/
GenBank, http://www.ncbi.nlm.nih.gov/GenBank/ (for *MUTYH* [accession number AF527839], *OGG1* [accession number NT_022517], and *MTH1* [accession number NT_007819])
GENSCAN, http://genes.mit.edu/GENSCAN.html
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *MYH, OGG1, MTH1, APC,* and colorectal cancer)
Pfam, http://www.sanger.ac.uk/Software/Pfam/
PolyPhen, http://www.bork.embl-heidelberg.de/PolyPhen/
SIFT, http://blocks.fhcrc.org/%7Epauline/SIFT
T-Coffee, http://www.ch.embnet.org/software/TCoffee.html

References

Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP (2002) Inherited variants of *MYH* associated with somatic G:C→T:A mutations in colorectal tumors. Nat Genet 30:227–232

Croitoru ME, Cleary SP, Di Nicola N, Manno M, Selander T, Aronson M, Redston M, Cotterchio M, Knight J, Gryfe R, Gallinger S (2004) Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. J Natl Cancer Inst 96:1631–1634

Enholm S, Hienonen T, Suomalainen A, Lipton L, Tomlinson I, Karja V, Eskelinen M, Mecklin JP, Karhu A, Jarvinen HJ, Aaltonen LA (2003) Proportion and phenotype of *MYH*-associated colorectal neoplasia in a population-based series of Finnish colorectal cancer patients. Am J Pathol 163:827–832

Fleischmann C, Peto J, Cheadle J, Shah B, Sampson J, Houlston RS (2004) Comprehensive analysis of the contribution of germline *MYH* variation to early-onset colorectal cancer. Int J Cancer 109:554–558

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411:599–603

Sampson JR, Dolwani S, Jones S, Eccles D, Ellis A, Evans DG, Frayling I, Jordan S, Maher ER, Mak T, Maynard J, Pigatto F, Shaw J, Cheadle JP (2003) Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of *MYH*. Lancet 362:39–41

Satagopan JM, Offit K, Foulkes W, Robson ME, Wacholder S, Eng CM, Karp SE, Begg CB (2001) The lifetime risks of breast cancer in Ashkenazi Jewish carriers of *BRCA1* and *BRCA2* mutations. Cancer Epidemiol Biomarkers Prev 10: 467–473

Sieber OM, Howarth KM, Thirlwell C, Rowan A, Mandir N, Goodlad RA, Gilkar A, Spencer-Dene B, Stamp G, Johnson V, Silver A, Yang H, Miller JH, Ilyas M, Tomlinson IP (2004) *Myh* deficiency enhances intestinal tumorigenesis in multiple intestinal neoplasia ($Apc^{Min/+}$) mice. Cancer Res 64:8876–8881

Sieber OM, Lipton L, Crabtree M, Heinimann K, Fidalgo P, Phillips RK, Bisgaard ML, Orntoft TF, Aaltonen LA, Hodgson SV, Thomas HJ, Tomlinson IP (2003) Multiple colorectal adenomas, classic adenomatous polyposis, and germline mutations in *MYH*. N Engl J Med 348:791–799

Wang L, Baudhuin LM, Boardman LA, Steenblock KJ, Petersen GM, Halling KC, French AJ, Johnson RA, Burgart LJ, Rabe K, Lindor NM, Thibodeau SN (2004) *MYH* mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps. Gastroenterology 127:9–16