Carcinogenesis vol.27 no.11 pp.2235–2242, 2006
doi:10.1093/carcin/bgl089
Advance Access publication June 13, 2006

Common variants in mismatch repair genes and risk of invasive ovarian cancer

Honglin Song[1,*], Susan J.Ramus[2], Lydia Quaye[2], Richard A.DiCioccio[3], Jonathan Tyrer[1], Emma Lomas[1], Danielle Shadforth[1], Estrid Hogdall[4], Claus Hogdall[5], Valerie McGuire[6], Alice S.Whittemore[6], Douglas F.Easton[7], Bruce A.J.Ponder[1], Susanne Kruger Kjaer[4], Paul D.P.Pharoah[1] and Simon A.Gayther[2]

[1]CR-UK Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK, [2]Translational Research Laboratory, Department of Gynaecological Oncology, University College London, UK, [3]Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY, USA, [4]Department of Viruses, Hormones, and Cancer, Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark, [5]Gynaecologic Clinic, The Juliane Marie Centre, Rigshospitalet, University of Copenhagen, Denmark, [6]Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA and [7]CR-UK Genetic Epidemiology Unit, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK

*To whom correspondence should be addressed. Tel: +0044 1223 740161, Fax: +0044 1223 740147;
Email: honglin@srl.cam.ac.uk

Mismatch repair (MMR) is important for repairing of nucleotide mismatches during DNA replication. Germline mutations in MMR genes are associated with hereditary non-polyposis colorectal cancer (HNPCC). Ovarian cancer occurs as part of the HNPCC phenotype, and so common variants in MMR genes are candidates for ovarian cancer susceptibility. We performed a large multicentre case–control study to investigate associations of common variations in MMR genes and ovarian cancer using a single nucleotide polymorphism (SNP) tagging approach. A total of 2570 controls and 1531 cases from three separate studies were genotyped for 44 tagging SNPs (stSNP) in seven MMR genes (*MLH1, MLH3, MSH2, MSH3, MSH6, PMS1* and *PMS2*). Genotype frequencies were marginally different between cases and controls for *PMS2* rs7797466 ($P_{2df} = 0.046$) with a 1.17-fold (95% CI 1.03–1.33) increase in risk for each '*a*' allele carried (*P*-trend = 0.013). Haplotype analysis of *PMS2* also showed significant differences in frequencies between cases and controls ($P_{7df} = 0.005$), with one haplotype accounting for most of the effect. There was also marginal evidence for a recessive protective effect with common homozygote as the baseline comparator for two SNPs—*MSH6* rs3136245 (OR 0.67; 95% CI 0.46–0.98) and *MSH3* rs6151662 (OR 0.28; 95% CI 0.08–0.91)—but the comparisons of genotype frequencies for these variants were not significant ($P = 0.10$ and 0.054). In conclusion, it is unlikely that common variants in *MLH1, MLH3, PMS1, MSH2, MSH3* and *MSH6* contribute significantly to ovarian cancer susceptibility. The observed association of *PMS2* rs7797466 with ovarian cancer warrants confirmation in an independent study.

Abbreviations: EGP, NIEHS Environmental Genome Project; HNPCC, hereditary non-polyposis colorectal cancer; LD, linkage disequilibrium; MAF, minor allele frequency; MMR, mismatch repair; MSI, microsatellite instability; stSNP, SNP tagging SNPs.

Introduction

Ovarian cancer is the sixth most common malignancy occurring in women and causes 125 000 deaths annually in women worldwide (1). The aetiology of ovarian cancer is not fully understood—but family and twin studies suggest that inherited factors are an important cause (2). The known ovarian cancer susceptibility genes, such as *BRCA1* and *BRCA2,* explain <30% of the excess familial risk of ovarian cancer. It has been hypothesized that common, low to moderate penetrance alleles account for most of the remaining risk.

It is widely accepted that human neoplastic transformation including ovarian cancer is a multi-step process and involves the accumulation of genetic alterations. Mismatch repair (MMR) is one of the most important DNA repair processes. It is responsible for the repair of nucleotide mismatches during DNA replication and prevents the propagation of potentially harmful mutations in genes including those involved in cancer development (3–5). Loss of MMR function leads to genetic instability, which can be identified as microsatellite instability (MSI) in tumours with deficient MMR. MSI has been identified in multiple different types of sporadic cancer, including prostate (6), pancreatic (7), gastric cancers (8) and ovarian cancer (9). Germline mutations in MMR genes are also responsible for the inherited susceptibility syndrome, hereditary non-polyposis colorectal cancer (HNPCC), which is typically characterized by cancer of the colorectum, but also cancers in the upper gastrointestinal tract, the endometrium and ovary (10). To date all but *MSH3* have been shown to be associated with an inherited predisposition to cancer. MMR gene mutation carriers have nearly a 10-fold rise in risk of ovarian cancer in female (3). Furthermore, hypermethylation of the *MLH1* promoter region has been found in MSI cancers of ovary (11) and endometrium (12) that do not contain germline *MLH1* mutations. Geisler *et al*. (13) reported that 16.8% of ovarian carcinoma cases have defects of the DNA MMR system. A recent study demonstrated that MMR deficiencies was involved in the development and/or progression of a proportion of epithelial ovarian cancers through the accumulation of genetic alterations (14).

The seven known genes in the MMR pathways—*MSH2, MSH3, MSH6, MLH1, MLH3, PMS1* and *PMS2*—are located on five different chromosomes. *MSH2, MSH6* and *PMS1* are on chromosome 2, *MLH1* is on chromosome 3, *MSH3* is on chromosome 5, *PMS2* is on chromosome 7 and *MLH3* is on chromosome 14. The coding and non-coding content of these genes ranges from ~24 kb (*MSH6*) to ~222 kb (*MSH3*). The

protein products have a variety of functions in the recognition and repair of DNA mismatches. MSH2/MSH6 (MutSα) and MSH2/MSH3 (MutSβ) form heterodimers that recognize and bind to the mismatch region (15); MutL (MLH1/PMS1 or MLH1/PMS2) is involved with mismatch strand excision and subsequent repair (15). Gene defects in *MSH2* result in the loss of both MutSα and MutSβ and complete inactivation of MMR function. MSH6 and MSH3 have some overlap in their function—consequently, inactivation of either of these genes individually has a lesser impact (16). *MLH3* was only recently identified and found to be associated with mammalian MSI (17).

Most high penetrance MMR mutations lead to loss of protein function and genetic instability and ultimately predispose to cancer. However, it is also possible that there are several common polymorphisms in these genes that may be associated with variation in cancer risk. Recently, the rare *MLH1* D132H variant was found to be associated with susceptibility to colorectal cancer (18) and polymorphisms in *PMS2* were associated with HNPCC (19). Because ovarian cancer occurs as part of the HNPCC phenotype, we hypothesize that common *MMR* gene variants may lead to variation in ovarian cancer risk. The purpose of this study was to investigate whether any associations exist between disease risk and *MMR* polymorphic variation in a large, multicentre ovarian cancer case–control study.

Materials and methods

Study subjects

The cases and controls used for this study were from three different case–control studies: the SEARCH study from the UK, the Malignant Ovarian Cancer (MALOVA) study from Denmark and the Family Registry for Ovarian Cancer (FROC) study from the USA (Table I).

The SEARCH ovarian cancer study is an ongoing, population-based ovarian cancer case–control study covering the regions served by the East Anglia and West Midlands cancer registries in the UK. All patients diagnosed in East Anglia with invasive epithelial ovarian cancer under the age of 70 years since 1991 and still alive in 1998 when recruitment started to take part (prevalent cases). Incident cases are those diagnosed <70 years since 1998 in East Anglia and since 2003 in the West Midlands. From 1991 to the end of 1997, 1181 women were registered in East Anglia of whom 767 had already died and the general practitioner refused permission to contact 166. Thus, we invited 248 women to take part, of whom 216 provided a blood sample (87% of those invited and 18% of all eligible diagnoses in the region). As the study is ongoing, the following data are based on registrations in East Anglia from 1998 to 2004 for which recruitment has been completed. In this period 1453 women were registered of whom 334 had died by the time of registration and the general practitioner refused permission to contact 531 (reason unknown). Of 588 women invited to take part 476 provided a blood sample (81% of those invited and 33% of all eligible diagnoses). To date we have invited 1750 women to participate of whom 1157 have provided a blood sample—the first 732 cases were available for this analysis. Female controls (n = 855) have been randomly selected from the EPIC-Norfolk component of the European Prospective Investigation of Cancer (EPIC), a prospective study of diet and cancer being carried out in East Anglia. The EPIC Norfolk cohort comprises 25 000 individuals resident in Norfolk, aged 35–74 years at first interview in 1993. Blood for DNA extraction was collected during the second health check in 1998–2000 (84%). The ethnic background of cases and controls is similar, with over 98% being white Europeans. Participants are asked to provide written consent, to complete an epidemiological questionnaire and to provide a 20 ml whole blood sample. The study is approved by the Eastern Multi-centre Research Ethics Committee. DNA was extracted from blood samples by Whatman International Ltd (Ely, UK) using a chloroform–phenol method.

Table I. Details of study subjects

Study	Cases		Controls	
	Number	Mean age[a] (range)	Number	Mean age[b] (range)
SEARCH	732	55.6 (21–74)	855	52.7 (39–77)
MALOVA	472	59.8 (32–80)	1286	56.8 (33–80)
FROC	327	50.8 (23–64)	429	48.2 (19–66)
Total	1531		2570	

[a]Age at diagnosis.
[b]Age at recruitment.

The MALOVA study is a population-based, Danish case–control study of ovarian cancer. Eligible cases were women aged 30–80 years, who were diagnosed with an ovarian tumour from December 1994 to May 1999. The study included 18 different hospitals from the municipalities of Copenhagen and Frederiksberg as well as the counties of Copenhagen, Frederiksborg, Roskilde, Western Sealand, Storstrøm, Funen, Southern Jutland and Northern Jutland. By the end of the study period, a total of 959 ovarian cancer patients were identified in this study area. Of these, 53 patients were considered too ill to participate, 45 women died before being contacted. Thus, 861 were invited to take part of whom 652 (76%) provided a blood sample. Controls were drawn from the general female population within the study area (aged 30–80 years) selected at random using the computerized Central Population Register. Of 3137 eligible controls 2116 (67.5%) enrolled—1564 completed a personal interview and 552 completed a telephone interview. Samples from 472 cases and 1286 controls were available for this study. DNA was extracted from blood using the QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA). This study has been approved by the scientific ethical committee in the study area, KF01-384/95 and all subject provided written consent.

The USA subjects were ascertained from six counties in Northern California as part of the FROC. Patients with epithelial ovarian cancer diagnosed between March 1, 1997 and July 31, 2001 were identified through the Greater Bay Area Cancer Registry operated by the Northern California Cancer Center as part of the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute. Rapid case ascertainment was used to identify cases within 1 month of diagnosis. Eligible patients were those diagnosed with invasive or low malignant potential epithelial ovarian cancer at ages 20–64 years who resided in six Bay Area counties (Alameda, Contra Costa, Marin, San Francisco, San Mateo or Santa Clara). Of the 915 eligible patients, 24 were not contacted because their physicians refused permission, 209 refused to participate, had died or could not be located, and 682 (75%) were interviewed. Of these, 579 (85% of those interviewed and 63% of total) provided a blood or mouthwash sample for DNA extraction.

Control women were identified through random-digit dial and were frequency-matched to cases on race/ethnicity and 5-year age group. Of 1041 controls randomly selected into the study, 746 (72%) completed the telephone screening interview, 162 (16%) refused to participate and 133 (13%) could not be located. Of 689 control women who met the eligibility criteria (i.e. no prior history of breast or ovarian cancer), 626 (91%) completed the detailed family history and epidemiology questionnaires and 585 (85%) provided a blood or mouthwash sample. In addition, 244 first degree relatives of cases without ovarian or breast cancer were recruited for the study.

Samples of heparinized, peripheral blood or buccal mouthwash rinse were shipped to Roswell Park Cancer Institute, Buffalo, NY. Genomic DNA was isolated from leucocytes of peripheral blood using the Puregene Kit (Gentra Systems, Minneapolis, MN). Genomic DNA was also isolated from exfoliated cells in buccal mouthwash rinses as described previously (20). Research was conducted with protocols approved by the Institutional Review Boards of Stanford University School of Medicine and Roswell Park Cancer Institute. This analysis is restricted to the 327 white and white non-Hispanic cases with invasive epithelial ovarian cancer, and the 429 age- and race/ethnicity-matched controls who had not undergone bilateral oophorectomy. The relative controls were not included in this analysis.

Tag SNP selection

We started this project before detailed information on common genetic variation in MMR genes was available. Initially we selected several common SNPs of varying frequencies from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP) and other public databases. Shortly after this, resequencing data from the NIEHS Environmental Genome Project (EGP) (http://egp.gs.washington.edu/) became available for *MSH6* and *PMS1*. Resequencing enable us to carry out a more systematic selection of SNP tagging SNPs (stSNPs). The EGP has been resequencing candidate genes for cancer across panels of individuals representative of US ethnicities. The original

panel (P1-PDR90) of 90 individuals consists of 24 European Americans, 24 African Americans, 12 Mexican Americans, 6 Native Americans and 24 Asian Americans, but the ethnic group identifiers are not available. It is known that there is greater genetic and haplotype diversity in individuals of African origin; so we have identified and excluded 28 of the samples with the greatest African ancestry in this population by comparing the genotypes of the PDR90 sample with genotypes for the same SNPs from the National Heart, Lung and Blood Institute Variation Discovery Resource Project African American panel (http://pga.gs.washington.edu/finished_genes.html). Data from the remaining 62 individuals were used to identify stSNPs. If a candidate gene has not been resequenced by NIEHS EGP we have used genotype data from the International Hap Map project (http://www.hapmap.org). HapMap does not attempt to comprehensively identify all SNPs but genotypes a selection of common SNPs within each gene in 30 parent–offspring trios. These samples were collected in 1980 from the US residents with northern and western European ancestry by the Centre d'Etude du Polymorphisme Humain (CEPH).

The aim of the SNP tagging was to identify a set of stSNPs that efficiently tags all the known common variants (minor allele frequency (MAF) >0.05) and is likely to tag most of the unknown common variants in the gene. The best measure of the extent to which one SNP tags another SNP is the pairwise correlation coefficient (r_p^2), since the loss in power incurred by using a maker SNP in place of the true causal SNP is directly related to this measure. Where a SNP is poorly correlated with any other single SNP, it may be efficiently tagged by a haplotype defined by multiple SNPs, thus reducing the number of tagging SNPs needed. The correlation between a multi-SNP haplotype and an individual SNP is r_s^2. We aimed to achieve a minimum correlation between each SNP and a haplotype of tagging SNPs (r_s^2) to be 0.8. As tagging SNP selection based on r_s^2 will be inefficient where there is excessive haplotype diversity, we divided the gene into haplotype blocks and selected the stSNPs for each block separately using the programme tagSNPs (21). SNPs that were originally selected were forced in as stSNPs. If assay design failed for a selected stSNP the tagging selection was repeated with forced exclusion of the failed SNP in order to select an alternative (where possible). It is possible to use a variety of formal definitions of haplotype blocks; but we simply used the graphical representations of the pattern of linkage disequilibrium (LD) based on D' and selected blocks such that the common haplotypes in each block accounted for at least 80% of all haplotypes observed using the Haploview program (22). The selection of stSNPs for each gene was performed as follows:

MLH3: Neither HapMap nor EGP data were available for this gene at the start of the project. Therefore, we genotyped two SNPs available for this gene as Assays-on-Demand from Applied Biosystems (http://www.appliedbiosystems.com). These two SNPs generate three haplotypes, which account for 99% of all the haplotypes. Subsequently, HapMap released data for 11 different SNPs. However, these SNPs are in very strong LD and again, there were only three common haplotypes explaining 98% of haplotype diversity. As the 11 SNPs could be tagged by just two SNPs ($r_p^2 > 0.93$) it is highly likely that the two SNPs we have genotyped will have tagged the gene efficiently.

MSH2: Data for 26 common SNPs covering 82 kb were available in HapMap (21-06-2005 HapMap public release). There was a single LD block with common haplotypes accounting for 86% of all the haplotypes. Seven tagging SNPs were chosen; 21 out of 26 SNPs were tagged by $r_p^2 > 0.8$. The remainder were tagged by multimarker haplotype combination with mean r_s^2 of 0.9 and a minimum $r_s^2 = 0.72$.

MSH3: HapMap (21-06-2005 HapMap public release) included data for 81 common SNPs in single LD blocks in which the common haplotypes accounted for 83% of all haplotypes. Ten tagging SNPs were selected. For the 81 common SNPs mean r_p^2 was 0.79 and 50 SNPs were tagged by a single marker SNP with $r_p^2 > 0.8$. The remainder were tagged by multimarker haplotype combinations with mean r_s^2 of 0.93 and minimum r_s^2 of 0.82.

MSH6: EGP resequencing data covered 84% of the gene and identified 42 common SNPs. Three SNPs were poorly correlated with any others. After excluding these there were two LD blocks in which the common haplotypes accounted for 76 and 89% of all haplotypes. However, the presence of substantial LD between blocks allowed all SNPs to contribute to tagging outside its own block. Assay design failed for a total of nine SNPs that were initially selected as tags and ultimately there were six tagging SNPs in block 1 and one in block 2. The mean r_p^2 was 0.76 with 26 tagged by a single marker with $r_p^2 > 0.8$. The others were tagged with mean $r_s^2 = 0.78$ and minimum r_s^2 of 0.64.

PMS1: EGP resequencing data covered 55% of the gene and 74 common SNPs were identified. Two SNPs were poorly correlated with any others. After excluding these there were three LD blocks in which the common haplotypes comprised 92, 83 and 84% of all haplotypes, respectively. Again, the presence of substantial LD between blocks allowed tagging SNPs to contribute to tagging outside its own block. Three stSNPs were selected for block 1, one stSNP for Block 2 and three stSNPs for Block 3. The mean r_p^2 was 0.73 with 30 SNPs tagged by a single marker with $r_p^2 > 0.8$. The others were tagged by multimarker haplotype combinations within each block with mean r_s^2 of 0.92 and minimum r_s^2 of 0.78. We also genotyped rs3762545 that was not identified by the EGP resequencing, and so our tagging efficiency is likely to have been even better.

MLH1: EGP resequencing data covered 30% of the gene and 15 common SNPs were identified. There was a single LD block in which common haplotypes comprised 95% of all the haplotypes. Four stSNPs were chosen. Thirteen SNPs were tagged by single marker with minimum $r_p^2 > 0.91$, the other two SNPs were tagged by multi-SNP haplotypes with minimum $r_s^2 = 0.83$.

PMS2: No resequencing data were available for PMS2, so we used HapMap data (1-03-2005 HapMap public release) for this gene which comprised nine common SNPs within a single LD block. The common haplotypes accounted for 84% of the total. Seven stSNPs were selected, but two of these failed assay design and there were no alternatives. The other seven SNPs were tagged with minimum $r_p^2 > 0.9$. The SNPs with failed assays were poorly tagged by the others with r_p^2 of 0.13 and 0.17, respectively.

The number of common SNPs and number of tagging SNPs identified in each MMR gene are described in Table II.

Table II. Number of common SNPs and number of SNP tagging SNPs identified in each gene

Gene	Gene size (kb)	Database	Region covered (kb)	Common variants	Density (kb/SNP)	No. of LD blocks	No. of stSNPs
MSH2	80.1	HapMap	82	27	3.2	1	7
MSH3	222.1	HapMap	241	81	3.0	1	10
MSH6	23.8	EGP	22.6	42	0.6	2	7
PMS1	93.1	EGP	58.3	74	0.8	3	7
PMS2	35.8	HapMap	35	9	4.4	1	5
MLH1	57.4	EGP	18	15	1.3	1	4
MLH3	34.8	HapMap	42	11	4.2	1	2

Genotyping

All samples were genotyped using the Taqman 7900HT Sequence Detection System according to manufacturer's instructions. Each assay was carried out using 10 ng DNA in a 5 μl reaction using TaqMan universal PCR master mix (Applied Biosystems), forward and reverse primers and FAM and VIC labelled probes designed by Applied Biosystems (ABI Assay-by-design). Primer and probe sequences and assay conditions used for each polymorphism analysed are available from the corresponding author on request. All assays were carried out in 384-well arrays with 12 duplicate samples in each plate for quality control. Assays where genotypes in duplicated samples did not show >95% concordance were discarded and replaced with alternative assays with the same tagging properties. Genotypes were determined using Allelic Discrimination Sequence Detection Software (Applied Biosystems, Warrington, UK). DNA samples that did not give a clear genotype result at the first attempt were not repeated. Hence, there are variations in the number of samples successfully genotyped for each polymorphism.

Statistics

The three studies were treated as separate strata in the analyses. Deviations of the genotype frequencies in the controls from those expected under Hardy–Weinberg equilibrium (HWE) were assessed by χ^2-tests (1 d.f.). The primary tests of association were the univariate analyses between each tagging SNP and ovarian cancer. Genotype frequencies in cases and controls were compared for each study separately using χ^2-tests for heterogeneity (2 d.f.). The data were then pooled and genotype frequencies were compared in cases and controls using unconditional logistic regression with terms for genotype and study and an appropriate likelihood ratio test. We tested for heterogeneity between study strata by comparing logistic regression models with and without a genotype–stratum interaction term using likelihood ratio tests. Genotypic specific risks with the common homozygote as the baseline comparator were estimated as odds ratios (OR) with

associated 95% confidence interval (95% CI) by unconditional logistic regression. We also tested for rare allele dose effect (assuming a multiplicative model) using χ^2-tests (1 d.f.) for each study separately and unconditional logistic regression for the pooled data.

In addition to the univariate analyses, we carried out global haplotype tests for each gene/haplotype block, because some SNPs were tagged by multimarker haplotypes. Haplotype frequencies and subject-specific expected haplotype indicators were calculated separately for each study using the programme TagSNPs, which implements an expectation-substitution approach to account for haplotype uncertainty given unphased genotype data (23). Subjects missing >50% genotype data in each *MMR* gene were excluded from haplotype analysis. We considered haplotypes with >2% frequency in at least one study to be 'common'. Rare haplotypes were pooled. We used unconditional logistic regression to test the global null hypothesis of no association between haplotype frequency and ovarian cancer, by comparing a model with multiplicative effects for each common haplotype (treating the most common haplotype as the referent) to the intercept-only model. Haplotype specific odds ratios were also estimated with their associated confidence intervals.

Results

We have genotyped 42 SNPs in 7 *MMR* genes in three ovarian cancer case–control studies (Table III). An additional two SNPs—*MLH1* rs1799977 and *MSH6* rs3136317—were genotyped for the SEARCH and FROC studies, but the genotype call rate for the MALOVA samples was unsatisfactory. Out of 130 genotype distributions in controls 9 distributions deviated from HWE ($P < 0.05$) compared with 7 expected by chance, but this occurred in only one of the studies for any given SNP apart from *MSH3* rs10079641, which deviated from HWE in both SEARCH ($P = 0.001$) and FROC ($P = 0.04$). These findings are unlikely to be owing to genotyping errors because no SNP deviated from HWE in both cases and controls and the discrimination of genotypes for these assays was good.

Table III. MMR polymorphisms and genotype distributions of combined data

Gene		Controls				Cases				MAF		
LD block	dbSNP	AA	Aa	aa	Total	AA	Aa	aa	Total	control	case	*P*-het[a]
MLH1	rs1800734	1224	638	89	1951	825	414	67	1306	0.21	0.21	0.64
	rs1540354	1286	590	71	1947	855	403	52	1310	0.19	0.19	0.69
	rs1799977[b]	624	477	123	1224	507	418	97	1022	0.30	0.30	0.66
	rs2286939	634	988	426	2048	398	663	265	1326	0.45	0.45	0.73
MSH6	rs3136245	1235	623	95	1953	852	422	44	1318	0.21	0.19	0.11
LD1	rs3136272	967	1135	311	2413	611	650	193	1454	0.36	0.36	0.41
	rs1800932	1353	564	78	1995	867	407	49	1323	0.18	0.19	0.45
	rs2348244	1420	460	55	1935	939	323	25	1287	0.15	0.15	0.13
	rs3136317[b]	866	353	40	1259	681	328	32	1041	0.17	0.19	0.19
	rs1800935	1037	810	175	2022	666	542	116	1324	0.29	0.29	0.85
LD2	rs2020911	740	893	308	1941	492	633	189	1314	0.39	0.39	0.33
MSH2	rs4952887	1607	336	19	1962	1103	187	17	1307	0.10	0.09	0.09
	rs13425206	1196	175	7	2478	1373	101	5	1479	0.04	0.04	0.85
	rs3771274	702	949	314	1965	469	633	207	1309	0.40	0.40	0.97
	rs1981928	1056	768	169	1993	708	518	96	1322	0.28	0.27	0.47
	rs3771281	783	925	301	2009	515	620	188	1323	0.38	0.38	0.88
	rs2059520	853	868	247	1968	567	587	163	1317	0.35	0.35	0.96
	rs2303428	1641	377	26	2044	1096	219	10	1325	0.11	0.09	0.17
MSH3	rs6151662	2173	258	16	2448	1315	160	3	1477	0.06	0.06	0.054
	rs40139	686	930	406	2022	414	669	241	1324	0.43	0.44	0.0504
	rs26282	1065	773	146	1984	697	520	99	1316	0.27	0.27	0.92
	rs26779	903	1161	391	2455	546	709	224	1479	0.40	0.39	0.87
	rs33008	1041	765	164	1970	680	539	97	1316	0.28	0.28	0.48
	rs10079641	2058	369	30	2457	1220	243	14	1477	0.09	0.09	0.31
	rs184967	1417	515	46	1978	931	349	33	1313	0.15	0.16	0.93
	rs2897298	1912	492	34	2438	1119	331	20	1470	0.12	0.13	0.32
	rs26279	1064	771	168	2003	663	525	122	1310	0.28	0.29	0.53
	rs2112416	1452	477	43	1972	975	308	32	1315	0.14	0.14	0.83
PMS1	rs3762545	1268	658	87	2013	857	411	48	1316	0.21	0.19	0.45
	rs5742981	2090	190	6	2286	1214	90	1	1305	0.04	0.04	0.48
LD1	rs5741593	2105	318	12	2435	1302	168	5	1475	0.07	0.06	0.48
	rs1233291	1007	811	146	1964	721	506	86	1313	0.28	0.26	0.23
LD2	rs1233255	1206	599	92	1897	809	421	49	1279	0.21	0.20	0.33
LD3	rs1233258	987	807	162	1956	697	509	104	1310	0.29	0.27	0.58
	rs256571	2176	288	8	2472	1296	180	2	1478	0.06	0.06	0.43
	rs256563	1875	541	34	2450	1164	289	16	1469	0.12	0.11	0.16
PMS2	rs7797466	1326	563	66	1956	842	415	54	1310	0.18	0.20	**0.046**
	rs2345060	1125	729	120	1974	750	508	66	1324	0.25	0.24	0.33
	rs2286680	1527	456	37	2020	980	318	27	1325	0.13	0.14	0.57
	rs12112229	1081	743	115	1939	728	491	89	1308	0.25	0.26	0.53
	rs1805321	636	982	349	1967	464	622	236	1322	0.43	0.41	0.17
	rs2228006	1812	612	41	2465	1066	372	33	1471	0.14	0.15	0.54
MLH3	rs7303	528	916	453	1897	339	621	328	1288	0.48	0.50	0.64
	rs175080	589	953	436	1978	408	639	257	1304	0.46	0.44	0.40

AA, common homozygote; Aa, heterozygote; aa, rare homozygote. Data in bold highlights the statistic significant results.
[a]Comparison of genotype frequencies in cases and controls (χ^2; 2 d.f.) in combined data.
[b]Genotype data missing for MALOVA study.

The genotype distributions and genotypic specific risks for each study are presented in Supplementary Table S1. The genotype frequencies in controls were similar for all three study populations ($P > 0.05$), apart from five SNPs in *MSH3* (rs6151662, rs40139, rs26779, rs2897298 and rs2112416; $P < 0.05$) (Supplementary Table S1). The observed genotype frequencies from the combined dataset are presented in Table III, which also shows the result of the test for heterogeneity of genotype frequencies (P-het) between cases and controls in combined data. The genotypic specific risks for the combined data and the result of the trend test for association are presented in Table IV. There was no difference in genotype frequency in ovarian cases and controls for 43 of SNPs tested. However, there was a marginally significant association for *PMS2* rs7797466 ($P = 0.046$) with a 1.17-fold (95% CI 1.03–1.3) increase in risk for each 'a' allele carried (P-trend $= 0.013$). There was no evidence of heterogeneity between studies ($P = 0.10$). There was some evidence for a recessive effect for two SNPs—rare allele homozygotes for *MSH6* rs3136245 and *MSH3* rs6151662 being at reduced risk of ovarian cancer [OR 95% CI 0.67 (0.46–0.97) and 0.28 (0.08–0.91), respectively]—but the comparisons of genotype frequencies for these variants were not significant ($P = 0.10$ and 0.054). There was no evidence of heterogeneity between studies for rs3136245 ($P = 0.57$) but with some evidence of heterogeneity between studies for rs6151662 ($P = 0.04$).

The haplotype frequencies for each gene in cases and controls in combined analysis were estimated after stratification and the results for haplotype specific risks and association test are presented in Table V. No significant differences in haplotype frequencies were seen between cases and controls for *MLH1, MLH3, PMS1, MSH2* and *MSH3*. However, for *PMS2* the haplotype frequencies differed significantly (global test $P = 0.005$). This was owing to the effect of a single haplotype comprising the rare allele of rs7797466 and the common allele of the other five SNPs [OR = 1.51 95% CI (1.18–1.94)]. One *MSH6* haplotype frequency was slightly different in cases and controls ($P = 0.04$) and had reduced risk [OR = 0.78; 95% CI (0.61–0.99)]; however, the global test is not significant ($P_{6df} = 0.31$).

Table IV. Genotypic specific risks (OR and 95% CI) of combined data

Gene	dbSNP	Heterologous OR[a] (95% CI)	Homologous OR[a] (95% CI)	*P*-trend
MLH1	rs1800734	0.94 (0.81–1.10)	1.08 (0.77–1.51)	0.78
	rs1540354	1.05 (0.90–1.22)	1.15 (0.79–1.67)	0.40
	rs1799977	1.08 (0.90–1.28)	0.97 (0.73–1.30)	0.78
	rs2286939	1.05 (0.89–1.23)	0.98 (0.80–1.20)	0.92
MSH6	rs3136245	0.97 (0.83–1.13)	**0.67 (0.46–0.97)**	0.13
	rs3136272	0.92 (0.80–1.06)	1.00 (0.82–1.24)	0.72
	rs1800932	1.10 (0.94–1.29)	0.98 (0.68–1.41)	0.40
	rs2348244	1.05 (0.89–1.24)	0.63 (0.39–1.03)	0.58
	rs3136317	1.18 (0.99–1.42)	1.02 (0.64–1.65)	0.15
	rs1800935	1.04 (0.90–1.21)	1.06 (0.81–1.36)	0.60
	rs2020911	1.04 (0.89–1.22)	0.89 (0.72–1.11)	0.50
MSH2	rs4952887	0.82 (0.67–0.99)	1.30 (0.66–2.54)	0.16
	rs13425206	0.93 (0.71–1.21)	1.05 (0.34–3.28)	0.63
	rs3771274	1.02 (0.87–1.19)	1.00 (0.80–1.24)	0.96
	rs1981928	1.02 (0.88–1.18)	0.86 (0.65–1.12)	0.52
	rs3771281	1.02 (0.88–1.19)	0.97 (0.78–1.20)	0.89
	rs2059520	1.02 (0.88–1.19)	1.00 (0.80–1.25)	0.92
	rs2303428	0.89 (0.73–1.07)	0.61 (0.31–1.21)	0.08
MSH3	rs6151662	1.01 (0.81–1.25)	0.28 (0.08–0.91)	0.40
	rs40139	1.17 (1.00–1.37)	0.96 (0.78–1.17)	0.97
	rs26282	1.03 (0.89–1.20)	1.01 (0.77–1.33)	0.76
	rs26779	1.01 (0.87–1.16)	0.95 (0.78–1.16)	0.73
	rs33008	1.09 (0.94–1.26)	0.91 (0.69–1.19)	0.95
	rs10079641	1.09 (0.91–1.31)	0.71 (0.39–1.31)	0.77
	rs184967	1.01 (0.87–1.19)	1.08 (0.68–1.72)	0.73
	rs2897298	1.11 (0.95–1.31)	0.83 (0.47–1.45)	0.43
	rs26279	1.08 (0.93–1.25)	1.10 (0.86–1.43)	0.27
	rs2112416	0.98 (0.83–1.15)	1.13 (0.71–1.80)	0.96
PMS1	rs3762545	0.93 (0.80–1.09)	0.83 (0.57–1.20)	0.16
	rs5742981	0.92 (0.70–1.20)	0.41 (0.06–2.61)	0.38
	rs5741593	0.89 (0.73–1.09)	0.76 (0.26–2.26)	0.23
	rs1233291	0.89 (0.77–1.03)	0.86 (0.65–1.15)	0.10
	rs1233255	1.04 (0.89–1.22)	0.79 (0.55–1.14)	0.70
	rs1233258	0.93 (0.80–1.08)	0.94 (0.72–1.23)	0.39
	rs256571	1.05 (0.86–1.28)	0.39 (0.07–2.06)	0.92
	rs256563	0.86 (0.73–1.02)	0.77 (0.42–1.39)	0.14
PMS2	rs7797466	**1.17 (1.01–1.38)**	1.37 (0.94–1.99)	**0.013**
	rs2345060	1.04 (0.90–1.20)	0.81 (0.59–1.12)	0.65
	rs2286680	1.09 (0.92–1.28)	1.13 (0.67–1.91)	0.30
	rs12112229	0.96 (0.83–1.11)	1.14 (0.85–1.55)	0.80
	rs1805321	0.86 (0.73–1.01)	0.93 (0.76–1.14)	0.29
	rs2228006	1.01 (0.87–1.18)	1.31 (0.81–2.12)	0.51
MLH3	rs7303	1.04 (0.88–1.23)	1.09 (0.90–1.34)	0.35
	rs175080	0.97 (0.82–1.14)	0.88 (0.72–1.08)	0.21

Data in bold highlights the statistic significant results.
[a]Compared with common homozygote.

Discussion

To our knowledge, this is the first comprehensive study to systematically tag all the known common variants in the genes in the MMR pathway and to test the tag SNPs for association with ovarian cancer susceptibility. We found borderline evidence for an association of *PMS2* rs7797466 (P-heterogeneity $= 0.046$, P-trend $= 0.013$) and haplotype analysis of *PMS2* showed significant association in a global test ($P_{7df} = 0.005$) with one rare haplotype (frequency 4%) accounting for most of the effect. *PMS2* rs7797466 encodes a variant in intron 1 of the gene which is not in a conserved region (http://genome.ucsc.edu/cgi-bin/hgGateway) and is unlikely to have a direct functional effect. It is correlated with two other SNPs ($r^2_p = 0.7$), both of which are intronic and unlikely to be functional. There are two rare coding SNPs in the gene, rs1805324 M622I (2.2%) and rs1805318 T597S (3.3%), but neither of these are correlated with rs7797466 ($r^2_p < 0.1$). The EGP is planning to resequence this gene at some time in the future. Such data may identify other putative functional variants tagged by the associated SNP/haplotype. We also found weaker evidence of disease associations for the SNPs *MSH6* rs3136245 and *MSH3* rs6151662. Both variants are intronic and not located within conserved regions of either gene, so are unlikely to have any functional effect; but they may be linked to as yet unidentified functional variants.

These results need to be interpreted with some caution. First, the comparison of genotype frequencies in ovarian cases and controls were marginally significant for only one SNP (*PMS2* rs7797466) out of 44 SNPs tested, and this result may have occurred by chance. The P-values presented above have not been adjusted for multiple hypothesis testing. Since SNPs within the same gene may be in linkage disequilibrium, the test statistics for the 44 SNPs are not independent, and standard methods for adjusting for multiple testing, such as the Bonferroni correction, would be too conservative. We

Table V. Estimated haplotype frequencies in cases and controls and haplotypic specific risks

Gene	Haplotype[a]	Control frequency	Case frequency	OR (95% CI)	χ^2 *P*-value	Global test *P*-value
MLH1[b]	h001	0.45	0.45	1.0		$\chi^2 = 1.63$
	H100	0.21	0.21	0.99 (0.86–1.12)	0.83	$P_{4df} = 0.80$
	h010	0.19	0.19	1.05 (0.91–1.20)	0.52	
	h000	0.15	0.15	0.99 (0.86–1.16)	0.94	
	Rare	0.002	0.001	0.44 (0.08–2.58)	0.36	
MSH6[c]	h000000	0.43	0.45	1.0		$\chi^2 = 7.11$
	h011011	0.18	0.19	0.99 (0.86–1.13)	0.85	$P_{6df} = 0.31$
	h100101	0.15	0.14	0.94 (0.80–1.09)	0.40	
	h010010	0.10	0.10	0.93 (0.79–1.11)	0.44	
	h010000	0.07	0.07	0.87 (0.71–1.08)	0.20	
	h100001	0.06	0.05	0.78 (0.61–0.99)	0.04	
	Rare	0.02	0.01	0.75 (0.49–1.15)	0.19	
MSH2	h0000000	0.50	0.51	1.0		$\chi^2 = 7.61$
	h0011110	0.17	0.17	1.0 (0.87–1.15)	0.99	$P_{8df} = 0.47$
	h0011111	0.10	0.09	0.86 (0.72–1.03)	0.11	
	h0010110	0.06	0.07	1.12 (0.91–1.37)	0.30	
	h1000000	0.05	0.04	0.84 (0.64–1.09)	0.18	
	h0100000	0.04	0.04	0.97 (0.74–1.26)	0.81	
	h1010100	0.03	0.03	1.0 (0.76–1.32)	0.99	
	h0010000	0.02	0.02	1.10 (0.76–1.61)	0.61	
	Rare	0.03	0.03	0.82 (0.58–1.15)	0.25	
MSH3	h0100100000	0.21	0.21	1.0		$\chi^2 = 1.51$
	h0001000000	0.16	0.15	0.95 (0.80–1.13)	0.56	$P_{10df} = 1.0$
	h0010001010	0.14	0.15	1.0 (0.84–1.17)	0.95	
	h0001000110	0.11	0.11	1.0 (0.84–1.21)	0.97	
	h0101010000	0.08	0.08	1.0 (0.82–1.22)	0.99	
	h0010000001	0.08	0.08	1.01 (0.82–1.26)	0.90	
	h1100100000	0.06	0.06	0.92 (0.73–1.16)	0.49	
	h0100000001	0.05	0.05	0.96 (0.74–1.24)	0.76	
	h0010000000	0.03	0.03	0.89 (0.64–1.24)	0.49	
	h0001000001	0.02	0.02	0.93 (0.61–1.41)	0.73	
	Rare	0.07	0.07	1.0 (0.80–1.25)	0.98	
PMS1	h0000000	0.64	0.65	1.0		$\chi^2 = 5.92$
	h0011100	0.10	0.10	1.04 (0.87–1.24)	0.69	$P_{7df} = 0.66$
	h0000100	0.07	0.07	1.03 (0.83–1.28)	0.80	
	h1010101	0.04	0.03	0.84 (0.64–1.09)	0.20	
	h0011000	0.04	0.04	0.90 (0.68–1.19)	0.47	
	h0111111	0.03	0.03	0.82 (0.61–1.10)	0.19	
	h0010101	0.02	0.01	0.85 (0.57–1.26)	0.42	
	h0111011	0.01	0.02	1.26 (0.83–1.91)	0.28	
	Rare	0.06	0.06	1.06 (0.84–1.33)	0.62	
PMS2	h010010	0.24	0.24	1.0		$\chi^2 = 20.29$
	h000010	0.18	0.17	0.96 (0.82–1.13)	0.61	**$P_{7df} = 0.005$**
	h000001	0.14	0.15	1.08 (0.91–1.27)	0.39	
	h000000	0.14	0.12	0.91 (0.76–1.09)	0.31	
	h101100	0.13	0.14	1.07 (0.90–1.26)	0.45	
	h000100	0.12	0.11	0.96 (0.80–1.15)	0.67	
	h100000	0.04	0.06	**1.51 (1.18–1.94)**	**0.001**	
	Rare	0.01	0.008	0.61 (0.33–1.10)	0.10	
MLH3	h10	0.48	0.50	1.0		$\chi^2 = 6.96$
	h01	0.46	0.45	0.94 (0.85–1.04)	0.27	$P_{3df} = 0.073$
	h00	0.06	0.06	0.92 (0.74–1.16)	0.49	
	Rare	0.003	0.00	0.09 (0.01–1.42)	0.09	

Data in bold highlights the statistic significant results.
[a]'0' represents common allele and '1' represents rare allele.
[b]Haplotype did not including rs1799977 because of MALOVA missing data.
[c]Haplotype did not including rs3136317 because of MALOVA missing data.

therefore used a simulation to determine an empirical *P*-value for the most significant result. In this analysis, we randomly shuffled the case–control status among individuals multiple times, and estimated how frequently a *P*-value <0.013 was obtained for from the randomly permuted data. In 1000 permutations $P < 0.013$ was observed on 426 occasions, giving *P*-value corrected for multiple testing of 0.43. This suggests that the marginally significant results that we observed may be the result of chance. There are a very large number of candidate ovarian cancer susceptibility polymorphisms and the prior probability of association is very low. Consequently, some authors have suggested that stringent criteria should be applied to statistical tests for genetic association, e.g. $P < 10^{-4}$ or even $P < 10^{-7}$ (24). A total of 7800 cases and an equal number of controls would be needed to detect a co-dominant allele with a minor allele frequency of 0.18 that confer a risk of 1.17 with 90% power and a Type I error rate of 10^{-4}.

Hidden population stratification is an alternative explanation for a spurious association. This occurs when allele frequencies differ between population subgroups and case and controls are drawn differentially from those subgroups.

It seems unlikely that population stratification is relevant in this association study because the cases and controls in the three studies reported here were drawn from the same ethnic groups. Furthermore, if stratification were present, it is unlikely that the same degree of stratification would be seen in all three studies. It is also worth noting that the existence of significant population stratification that has resulted in a false genetic association has never been empirically demonstrated (25). Finally, delay between diagnosis and recruitment into the studies may bias an association if a SNP is associated with prognosis. This is particularly relevant for the SEARCH study where a substantial proportion of potentially eligible cases had either died or permission to contact was refused. Nevertheless, there was no association between the time to blood draw after diagnosis for all but one of the SNPs ($P > 0.04$). For these SNPs, substantial survival bias is unlikely. *MSH3* rs26279 was associated with time to blood draw ($P = 0.006$), but this SNP was not associated with disease.

We found no evidence for association for the other 41 SNPs tested here with invasive ovarian cancer. This study include a minimum of 1279 cases and 1897 controls providing us with at least 89% power at the 0.001 level of significance to detect a co-dominant allele with frequency 0.3 that confers a relative risk of 1.3 or 76% power to detect a dominant allele with frequency of 0.1 that confers a relative risk of 1.3 with a type I error rate of 0.05. However, we cannot exclude the possibility that the alleles investigated are associated with smaller risks. Alternatively, there may be other susceptibility variants in these genes that are not strongly correlated with the polymorphisms examined. The SNPs tested were chosen in order to efficiently tag all the common variants in the genes, but assay design failed for eight stSNPs (four in *MSH6*, two in *PMS2* and one each in *MLH1* and *PMS1*), and alternative stSNPs were either not available or also failed assay design. One of the failed stSNPs (*PMS2* rs1059060) is a particularly good candidate because it encodes a non-synonymous amino acid change, Asn775Ser, that is situated in the PMS2–MLH1 interaction region (26). Furthermore, the data on which our stSNP selection was based are incomplete—for the genes where resequencing data were available, only 30–84% of the gene had been resequenced, so SNPs in the non-sequenced regions may not be tagged. In addition, rarer SNPs may have been missed by chance because the resequencing sample size was small (we based the stSNP selection on 62 individuals). It is not known how well stSNPs selected using HapMap data tag all the common variants in a gene, but where the HapMap SNP density high it is likely that most of the common variation is captured. We used HapMap data to tag *MSH3* but by the time the project was complete, EGP resequencing data were also available. Of the 201 common SNPs in the EGP data only 37 SNPs were also on HapMap, and just 4 of the 10 selected tag SNPs were identified in the EGP data. These 4 SNPs tagged 88 of the 201 with $r_p^2 > 0.8$ and mean r_p^2 of 0.55, but a much better performance would be expected for the full set of 10 stSNPs.

The absence of an association for common variants in MMR genes is not inconsistent with the fact that rare, deleterious mutation in these genes are associated with ovarian cancer as part of the HNPCC phenotype. Common variants in other high penetrance cancer susceptibility genes have been widely studied and none have emerged as definitive low-penetrance genes.

In conclusion, we have genotyped 44 SNP to tag the common variants in seven MMR genes in three studies. We found some evidence of association for *PMS2* rs7797466 and for a rare *PMS2* haplotype (global test $P_{7df} = 0.005$). We found little evidence of association for the other six MMR genes. It is unlikely that common variants in *MLH1, MLH3, MSH2, MSH3, MSH6* and *PMS1* contribute significantly to ovarian susceptibility, although we cannot exclude completely the possibility that other as yet unidentified variants are important (e.g. SNPs in distant regulatory regions or SNPs on haplotypes not efficiently tagged by those stSNPs genotyped to date). The observed association of *PMS2* rs7797466 with ovarian cancer risk warrants confirmation in an independent study and, if replicated, further investigation will be needed to identify the true functional variant.

Supplementary material

Supplementary data are available at *Carcinogenesis* online.

Acknowledgements

We thank Joan MacIntosh, Hannah Munday, Barbara Perkins, Clare Jordan, Kristy Driver, Mitul Shah, the local general practitioners and the East Anglian Cancer Registry for recruitment of the UK cases; the EPIC-Norfolk investigators for recruitment of the UK controls; Craig Luccarini and Don Conroy for expert technical assistance; Professor Ian Jacobs, Dr Usha Menon and Dr Jan Blaker for help in study design, fundraising and samples collection in Denmark; and all the study participants who contributed to this research. This work was funded by grants from Cancer Research UK, the Roswell Park Alliance, and the National Cancer Institute (CA71766 and Core Grant CA16056). H.S. was funded by a grant from WellBeing; S.J.R. is funded by a grant from the Eve Appeal; L.Q. is funded by a Special Trustees Foundation Grant from UCL; S.G. is a HEFC funded Senior Lecturer; B..A.J.P. is a Gibb Fellow, P.D.P.P. is a Senior Clinical Research Fellow and D.F.E. is a Principal Research Fellow of Cancer Research UK.

Conflict of Interest Statement: None declared.

References

1. Parkin,D.M., Bray,F., Ferlay,J. and Pisani,P. (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.
2. Pharoah,P.D. and Ponder,B.A. (2002) The genetics of ovarian cancer. *Best Pract. Res. Clin. Obstet. Gynaecol.*, **16**, 449–468.
3. Drake,A.C., Campbell,H., Porteous,M.E. and Dunlop,M.G. (2003) The contribution of DNA mismatch repair gene defects to the burden of gynecological cancer. *Int. J. Gynecol. Cancer*, **13**, 262–277.
4. Peltomaki,P. (1997) DNA mismatch repair gene mutations in human cancer. *Environ. Health Perspect.*, **105 Suppl 4**, 775–780.
5. Peltomaki,P. (2003) Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J. Clin. Oncol.*, **21**, 1174–1179.
6. Chen,Y., Wang,J., Fraig,M.M., Henderson,K., Bissada,N.K., Watson,D.K. and Schweinfest,C.W. (2003) Alterations in PMS2, MSH2 and MLH1 expression in human prostate cancer. *Int. J. Oncol.*, **22**, 1033–1043.
7. Park,J.G., Park,Y.J., Wijnen,J.T. and Vasen,H.F. (1999) Gene-environment interaction in hereditary nonpolyposis colorectal cancer with implications for diagnosis and genetic testing. *Int. J. Cancer.*, **82**, 516–519.
8. Yamamoto,H., Perez-Piteira,J., Yoshida,T., Terada,M., Itoh,F., Imai,K. and Perucho,M. (1999) Gastric cancers of the microsatellite mutator phenotype display characteristic genetic and clinical features. *Gastroenterology*, **116**, 1348–1357.
9. Arzimanoglou,I.I., Lallas,T., Osborne,M., Barber,H. and Gilbert,F. (1996) Microsatellite instability differences between familial and sporadic ovarian cancers. *Carcinogenesis*, **17**, 1799–1804.
10. Buermeyer,A.B., Deschenes,S.M., Baker,S.M. and Liskay,R.M. (1999) Mammalian DNA mismatch repair. *Annu. Rev. Genet.*, **33**, 533–564.

11. Gras,E., Catasus,L., Arguelles,R., Moreno-Bueno,G., Palacios,J., Gamallo,C., Matias-Guiu,X. and Prat,J. (2001) Microsatellite instability, MLH-1 promoter hypermethylation, and frameshift mutations at coding mononucleotide repeat microsatellites in ovarian tumors. *Cancer*, **92**, 2829–2836.
12. Simpkins,S.B., Bocker,T., Swisher,E.M., Mutch,D.G., Gersell,D.J., Kovatich,A.J., Palazzo,J.P., Fishel,R. and Goodfellow,P.J. (1999) MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. *Hum. Mol. Genet.*, **8**, 661–666.
13. Geisler,J.P., Goodheart,M.J., Sood,A.K., Holmes,R.J., Hatterman-Zogg,M.A. and Buller,R.E. (2003) Mismatch repair gene expression defects contribute to microsatellite instability in ovarian carcinoma. *Cancer*, **98**, 2199–2206.
14. Suzuki,M., Ohwada,M., Saga,Y., Saito,S. and Sato,I. (2001) Are DNA mismatch repair deficiencies responsible for accumulation of genetic alterations in epithelial ovarian cancers? *Cancer Genet. Cytogenet.*, **124**, 152–158.
15. Christmann,M., Tomicic,M.T., Roos,W.P. and Kaina,B. (2003) Mechanisms of human DNA repair: an update. *Toxicology*, **193**, 3–34.
16. Umar,A., Risinger,J.I., Glaab,W.E., Tindall,K.R., Barrett,J.C. and Kunkel,T.A. (1998) Functional overlap in mismatch repair by human MSH3 and MSH6. *Genetics*, **148**, 1637–1646.
17. Lipkin,S.M., Wang,V., Jacoby,R., Banerjee-Basu,S., Baxevanis,A.D., Lynch,H.T., Elliott,R.M. and Collins,F.S. (2000) MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat. Genet.*, **24**, 27–35.
18. Lipkin,S.M., Rozek,L.S., Rennert,G. *et al.* (2004) The MLH1 D132H variant is associated with susceptibility to sporadic colorectal cancer. *Nat. Genet.*, **36**, 694–699.
19. Yuan,Z.Q., Gottlieb,B., Beitel,L.K., Wong,N., Gordon,P.H., Wang,Q., Puisieux,A., Foulkes,W.D. and Trifiro,M. (2002) Polymorphisms and HNPCC: PMS2–MLH1 protein interactions diminished by single nucleotide polymorphisms. *Hum. Mutat.*, **19**, 108–113.
20. Lum,A. and Le Marchand,L. (1998) A simple mouthwash method for obtaining genomic DNA in molecular epidemiological studies. *Cancer Epidemiol. Biomarkers Prev.*, **7**, 719–724.
21. Stram,D.O., Haiman,C.A., Hirschhorn,J.N., Altshuler,D., Kolonel,L.N., Henderson,B.E. and Pike,M.C. (2003) Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. *Hum. Hered.*, **55**, 27–36.
22. Barrett,J.C., Fry,B., Maller,J. and Daly,M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.*, **21**, 263–265.
23. Zaykin,D.V., Westfall,P.H., Young,S.S., Karnoub,M.A., Wagner,M.J. and Ehm,M.G. (2002) Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum. Hered.*, **53**, 79–91.
24. Freimer,N.B. and Sabatti,C. (2005) Guidelines for association studies in Human Molecular Genetics. *Hum. Mol. Genet.*, **14**, 2481–2483.
25. Cardon,L.R. and Palmer,L.J. (2003) Population stratification and spurious allelic association. *Lancet*, **361**, 598–604.
26. Guerrette,S., Acharya,S. and Fishel,R. (1999) The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J. Biol. Chem.*, **274**, 6336–6341.

Received January 10, 2006; revised May 9, 2006; accepted May 9, 2006