

Candidate Gene Association Study of Esophageal Squamous Cell Carcinoma in a High-Risk Region in Iran

Mohammad R. Akbari,^{1,2,3} Reza Malekzadeh,¹ Ramin Shakeri,¹ Dariush Nasrollahzadeh,¹ Maryam Foumani,² Yulong Sun,² Akram Pourshams,¹ Alireza Sadjadi,¹ Elham Jafari,¹ Masoud Sotoudeh,¹ Farin Kamangar,⁴ Paolo Boffetta,⁵ Sanford M. Dawsey,⁴ Parviz Ghadirian,⁶ and Steven A. Narod²

¹Digestive Disease Research Center, Tehran University of Medical Sciences, Tehran, Iran; ²Women's College Research Institute and ³Institute of Medical Science, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; ⁴Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; ⁵IARC, Lyon, France; and ⁶Epidemiology Research Unit Research Centre, CHUM-Hôtel-Dieu, University of Montreal, Montreal, Quebec, Canada

Abstract

There is a region with a high risk for esophageal squamous cell carcinoma (ESCC) in the northeast of Iran. Previous studies suggest that hereditary factors play a role in the high incidence of cancer in the region. We selected 22 functional variants (and 130 related tagSNPs) from 15 genes that have been associated previously with the risk of ESCC. We genotyped a primary set of samples from 451 Turkmens (197 cases and 254 controls). Seven of 152 variants were associated with ESCC at the $P = 0.05$ level; these single nucleotide polymorphisms were then studied in a validation set of 549 cases and 1,119 controls, which included both Turkmens and non-Turkmens. The association observed for a functional variant in ADH1B was confirmed in the validation set, and that of a tagSNP in MGMT, the association was borderline significant in the validation set, after correcting for multiple testing. The other 5 variants that were associated in the primary set were not significantly associated in the validation set. The histidine allele at codon 48 of ADH1B gene was associated with a significantly decreased risk of ESCC in the joint data set (primary and validation set) under a recessive model (odds ratio, 0.41; 95% confidence interval, 0.29-0.76; $P = 4 \times 10^{-4}$). The A allele of the rs7087131 variant of MGMT gene was associated with a decreased risk of ESCC under a dominant model (odds ratio, 0.79; 95% confidence interval, 0.64-0.96; $P = 0.02$). These results support the hypothesis that genetic predisposition plays a role in the high incidence of ESCC in Iran. [Cancer Res 2009;69(20):7994-8000]

Introduction

Worldwide, 462,000 new cases of esophageal cancer are diagnosed annually (1). Seventy-five percent of affected individuals die within a year of diagnosis and the 5-year survival rate is only 10% to 16% (1). The incidence of esophageal cancer varies greatly between populations and across geographic regions. The east part of Golestan Province in northeastern Iran is a region with a very high risk of esophageal cancer, but the causative factors remain

elusive. The annual incidence rate exceeds 100/100,000 (2, 3), although the rate has declined over the last three decades (4). Turkmens constitute half of the population in this region and Persians, Turks, Sistanies, Balouches, and Kurds constitute the other half (referred to here as non-Turkmens).

Esophageal cancer is of two main histologic types: squamous cell carcinoma and adenocarcinoma. Esophageal squamous cell carcinomas (ESCC) constitute >90% of esophageal cancers among both Turkmens and non-Turkmens in eastern Golestan (5). We have recently shown that, among Turkmens, there is a strong familial component to esophageal cancer (6). The first gene found to contribute to the burden of esophageal cancer in Iran was BRCA2 (7). Protein-truncating mutations in BRCA2 are associated with an increased risk of esophageal cancer among Turkmens. However, because only 8% of the cases in the region carry a deleterious BRCA2 mutation (7), it is likely that other genes are involved.

Turkmens are believed to descend from Turkic tribes who migrated from the Altai Mountains (on the border of China and Mongolia) to northern Iran (8). A population at high risk for ESCC is also present in China (9) and the two high-risk populations might share common susceptibility alleles. Based on this hypothesis, we propose that the genes that have been reported previously to be associated with ESCC in the high-risk Chinese population (10, 11) are candidate genes among Turkmens. Initially, we focused on Turkmens because this is the largest ethnic group in the region and they are believed to be genetically homogeneous. However, we also studied a similar number of non-Turkmens from the same region. We studied 15 different genes that are involved in DNA repair, cell cycle control, and the metabolism of alcohol, folate, and carcinogens.

Materials and Methods

Study Subject

This report is a component of the Gastric and Esophageal Malignancies in Northern Iran Project (5, 12, 13), which is under way in Gonbad, the second largest city of Golestan Province, in northeastern Iran and in Ardabil, the largest city in Ardabil Province in northwestern Iran. The project is directed by the Digestive Disease Research Center of the Tehran University of Medical Sciences in collaboration with the National Cancer Institute, the IARC, and the University of Toronto. Cases and controls were collected between August 1, 2001 and May 15, 2008. The diagnosis of ESCC was confirmed for all cases by upper gastrointestinal endoscopy and by pathologic evaluation of tumor biopsies. The study protocol was approved by the ethics committees of the Digestive Disease Research Center, the National Cancer Institute, and the IARC.

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Requests for reprints: Steven A. Narod, Women's College Research Institute, Faculty of Medicine, University of Toronto, 7th Floor, 790 Bay Street, Toronto, Ontario, Canada M5G 1N8. Phone: 416-351-3765; Fax: 416-351-3767; E-mail: steven.narod@wchospital.ca

©2009 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-09-1149

We employed a two-stage design. In the first stage, we genotyped a total of 152 variants in 197 Turkmen cases and 254 Turkmen controls (the primary sample set). In the second stage, we genotyped the most promising variants identified from the primary sample set in 549 more cases and 1,119 more controls, including both Turkmens and non-Turkmens (the validation sample). The entire set of samples genotyped in this study is referred to as the joint sample set.

Cases. A total of 746 cases was enrolled in this study. There were 281 Turkmens and 220 non-Turkmens ESCC cases from Gonbad City. Non-Turkmen cases included Persians ($n = 114$), Turks ($n = 59$), Sistanies ($n = 28$), Balouches ($n = 9$), and Kurds ($n = 10$). In addition, we enrolled 245 Turkish cases from Ardabil, a city in northwestern Iran with intermediate rates of ESCC (13). The mean age at diagnosis was 63.6 years (range, 25-89 years); 380 cases (50.9%) were male and 366 cases (49.1%) were female. Epidemiologic data included smoking status, use of opium and alcohol, and a family history of cancer. The primary set of cases consisted of the first 197 Turkmens who were available for testing. The remaining 549 cases composed the validation set (84 Turkmens and 465 non-Turkmens).

Controls. One thousand three hundred seventy-three controls were recruited to the study: 244 were Turks from Ardabil City, 811 were Turkmen from Gonbad City, and 318 were non-Turkmen from Gonbad City [Persians ($n = 133$), Turks ($n = 78$), Sistanies ($n = 79$), Balouches ($n = 19$), and Kurds ($n = 9$)]. Eight hundred ninety-eight controls were patients referred to local hospitals for a reason other than cancer and 475 controls were healthy individuals enrolled from the Golestan cohort study (12). Controls had no personal history of cancer. The mean age for controls was 55.2 years (range, 24-90 years); 700 (51.0%) were male and 673 (49.0%) were female. The primary set of controls consisted of the first 254 Turkmens who were available for testing. The remaining 1,119 controls composed the validation set (557 Turkmens and 562 non-Turkmens). Characteristics of the total 746 cases and 1,373 controls are presented in Table 1.

Candidate Genes

Genes were selected based on literature review of the previously reported ESCC-associated variants, which mostly were from Chinese population. Four genes were involved in DNA repair ERCC2 (XPD), XRCC1, hOGG1, and MGMT; four genes were involved in cell cycle control: CCND1 (cyclin D1), CDKN2A (p16), TP53 (p53), and TMPRSS11A (ECRG1); four genes coded for carcinogen-metabolizing enzymes: CYP1A1, CYP2E1, GSTP1, and NAT2; two genes coded for alcohol-metabolizing enzymes: ADH1B and ALDH2; and

one gene coded for a folate-metabolizing enzyme: MTHFR. The role of these genes and their variants in esophageal cancer has been reviewed (10, 11).

For each gene, at least one functional variant has been reported in the literature to be associated with ESCC (10, 11, 14). These variants include Asp³¹²Asn (rs1799793) and Lys⁷⁵¹Gln (rs13181) from ERCC2 gene; Arg³⁹⁹Gln (rs25487) from XRCC1 gene; Ser³²⁶Cys (rs1052133) from hOGG1 gene; Leu⁷⁴Phe (rs12917) and Lys¹⁷⁸Arg (rs2308327) from MGMT gene; c.870A>G (rs9344) from CCND1 gene; Ala¹⁴⁸Thr (rs3731249) from CDKN2A gene; Arg⁷²Pro (rs1042522) from TP53 gene; Arg²⁹³Gln (rs353163) from TMPRSS11A (ECRG1) gene; Ile⁴⁶²Val (rs1048943) from CYP1A1 gene; c.-1019C>T (rs2031920) from CYP2E1; Ile¹⁰⁵Val (rs1695) from GSTP1 gene; Tyr⁹⁴Tyr (rs1041983), Ile¹¹⁴Thr (rs1801280), Leu¹⁶¹Leu (rs1799929), Arg¹⁹⁷Gln (rs1799930), and Gly²⁸⁶Glu (rs1799931) from NAT2 gene; Arg⁴⁸His (rs1229984) from ADH1B gene; Glu⁴⁸⁷Lys (rs671) from ALDH2 gene; and Ala²²²Val (rs1801133) and Glu⁴²⁹Ala (rs1801131) from MTHFR gene.

Three slow-metabolizing alleles of M1, M2, and M3 were reported for NAT2 gene. Variants Ile¹¹⁴Thr and Leu¹⁶¹Leu constitute allele M1, variants Arg¹⁹⁷Gln and Tyr⁹⁴Tyr constitute allele M2, and variant Gly²⁸⁶Glu is named allele M3 (15).

tagSNP Selection

Within populations, it is believed that single nucleotide polymorphisms (SNP) are distributed along chromosomes in haplotype blocks; therefore, genotyping of one or more SNPs, which are characteristic of a haplotype block, will capture the important information for the additional SNPs in that block. These SNPs are in linkage disequilibrium with the genotyped SNPs and are called tagSNPs. Genotype data from the Chinese population of the HapMap project (ref. 16; Rel21/phase II) were used for selecting tagSNPs to cover the genomic region of each gene, including 5 kbp in each 5' and 3' ends to cover the regulatory regions of the genes. Haploview software (17) was used for selecting tagSNPs. Aggressive tagging with using two- and three-marker haplotypes, LOD score threshold of 3.0 for multimarker testing, and r^2 threshold of 0.8 were employed for tagging procedure. Genomic regions of all 15 candidate genes encompass 705 kbp of the human genome. There are 635 SNPs with minor allele frequency of >1% in the Chinese population of the HapMap project. One hundred forty-one tagSNP were selected to capture 594 SNPs (94% of the total common SNPs) with mean r^2 of 0.98 in 152 tests (Table 2).

SNP Genotyping

We initially tested the 152 SNPs in a primary set of 197 cases and 254 controls. From the 141 tagSNPs selected for study, 11 were among the functional variants mentioned above and 130 were additional. All SNPs were genotyped in germ-line DNA of all samples of the primary set of cases and controls. iPLEX chemistry on a matrix-assisted laser desorption/ionization—time-of-flight MassARRAY System (Sequenom) was used for genotyping the 152 SNPs in eight reactions. The procedures were done according to the manufacturer's standard protocol. SNPs that showed association with ESCC in the primary set of samples ($P < 0.05$) were combined and a new assay design was done to genotype them in a single reaction. ESCC-associated SNPs identified in the primary set of samples were genotyped in the germ-line DNA of the validation set of samples at the second stage. Ten percent blinded quality-control samples were included in each plate; the concordance rates ranged from 98% to 100%. The average genotyping call rate was 98% (ranging from 95% to 100%). Two ESCC-associated SNPs were failed for genotyping in the primary set of samples; they were genotyped using TaqMan assay on ABI 7500 fast real-time system (Applied Biosystems). Genotyping of the primary set and validation set were done in Analytical Genetics Technology Centre of the University of Toronto and Genome Quebec Innovation Centre, respectively.

Statistical Analysis

The permutation version of the exact test was done to test for Hardy-Weinberg equilibrium. All genotypes were first studied in the primary sample set. Based on the results of the primary set, variants were selected for genotyping in the validation set. Haplotype analysis using expectation maximization algorithm (18) was used for comparing SNPs tagged by multiple markers. The significance level of $\alpha = 0.05$ was used for all

Table 1. Subject characteristics of the ESCC cases and controls

| Variable | Cases | Controls | P |
|-------------------------------|--------------|--------------|--------|
| Total <i>n</i> | 746 | 1,373 | — |
| Age, mean (range) | 63.6 (25-89) | 55.2 (24-90) | <0.001 |
| Gender, <i>n</i> (%) | | | |
| Male | 380 (50.9) | 700 (51.0) | 1.0 |
| Female | 366 (49.1) | 673 (49.0) | |
| Ethnicity, <i>n</i> (%) | | | |
| Turkmen | 281 (37.7) | 811 (59.0) | <0.001 |
| Turk | 304 (40.7) | 322 (23.5) | |
| Persian | 114 (15.3) | 133 (9.7) | |
| Sistani | 28 (3.7) | 79 (5.8) | |
| Balouch | 9 (1.2) | 19 (1.4) | |
| Kurd | 10 (1.3) | 9 (0.6) | |
| Ever smoker, <i>n</i> (%) | | | |
| Yes | 200 (27.9) | 339 (25.0) | 0.14 |
| No | 516 (72.1) | 1018 (75.0) | |
| Ever opium user, <i>n</i> (%) | | | |
| Yes | 158 (29.1) | 239 (20.2) | <0.001 |
| No | 382 (70.9) | 943 (79.8) | |

Table 2. Number of tagSNPs genotyped in this study along with the number of the SNPs captured with these tagSNPs for each gene

| No. | Gene | Chromosome | Size (kbp)* | Genotyped tagSNPs, <i>n</i> | Captured SNPs, <i>n</i> (%) [†] | Mean max <i>r</i> ² |
|-------|--------|------------|-------------|-----------------------------|--|--------------------------------|
| 1 | ADH1B | 4 | 10 | 3 | 4/4 (100) | 1.00 |
| 2 | ALDH2 | 12 | 53 | 10 | 22/23 (95) | 0.95 |
| 3 | CCND1 | 11 | 23 | 2 | 2/2 (100) | 1.00 |
| 4 | CDKN2A | 9 | 18 | 6 | 8/8 (100) | 0.99 |
| 5 | CYP1A1 | 15 | 16 | 1 | 1/1 (100) | 1.00 |
| 6 | CYP2E1 | 10 | 22 | 6 | 27/34 (79) | 0.97 |
| 7 | ECRG1 | 4 | 63 | 9 | 49/54 (90) | 0.99 |
| 8 | ERCC2 | 19 | 29 | 8 | 18/20 (90) | 0.97 |
| 9 | GSTP1 | 11 | 13 | 4 | 12/12 (100) | 0.99 |
| 10 | MGMT | 10 | 310 | 49 | 321/332 (96) | 0.95 |
| 11 | MTHFR | 1 | 30 | 8 | 35/36 (97) | 0.99 |
| 12 | NAT2 | 8 | 20 | 10 | 40/46 (86) | 0.94 |
| 13 | OGG1 | 3 | 27 | 5 | 5/5 (100) | 1.00 |
| 14 | TP53 | 17 | 29 | 6 | 7/14 (50) | 1.00 |
| 15 | XRCC1 | 19 | 42 | 14 | 43/44 (97) | 0.93 |
| Total | | | 705 | 141 | 594/635 (94) | 0.98 |

*Size of the gene, including 5 kbp, at each end for covering gene regulatory regions.

[†]The number of the SNPs with minor allele frequency >1% identified in the Chinese population of the HapMap project and the proportion of these SNPs that is captured by the genotyped tagSNPs in this study are shown for each gene.

comparisons in the primary sample set (the *P* values of the comparisons from the primary set were not adjusted for multiple testing). Because the direction of association in the validation set was based on the primary set results, a one-sided test was used for the comparisons in the validation set. In the validation set, the *P* values of the comparisons were corrected for multiple testing using the Bonferroni approach. For those variants that were independently confirmed in the validation set, the genotypes of the primary set and the validation set were combined to form a joint sample set. All

case-control comparisons were adjusted for age and ethnicity using multivariate logistic regression. Ethnicity was defined as Turkmen, Turk, Persian, Sistani, Balouch, or Kurd.

To exclude the possibility that there were technical errors in genotyping between the typing of subjects in the primary and validation sets, we re-genotyped all the 7 variants genotyped in the validation set in the primary sample set using the genotyping assay used for the validation sample set. The concordance rate of the genotyping results was 97%.

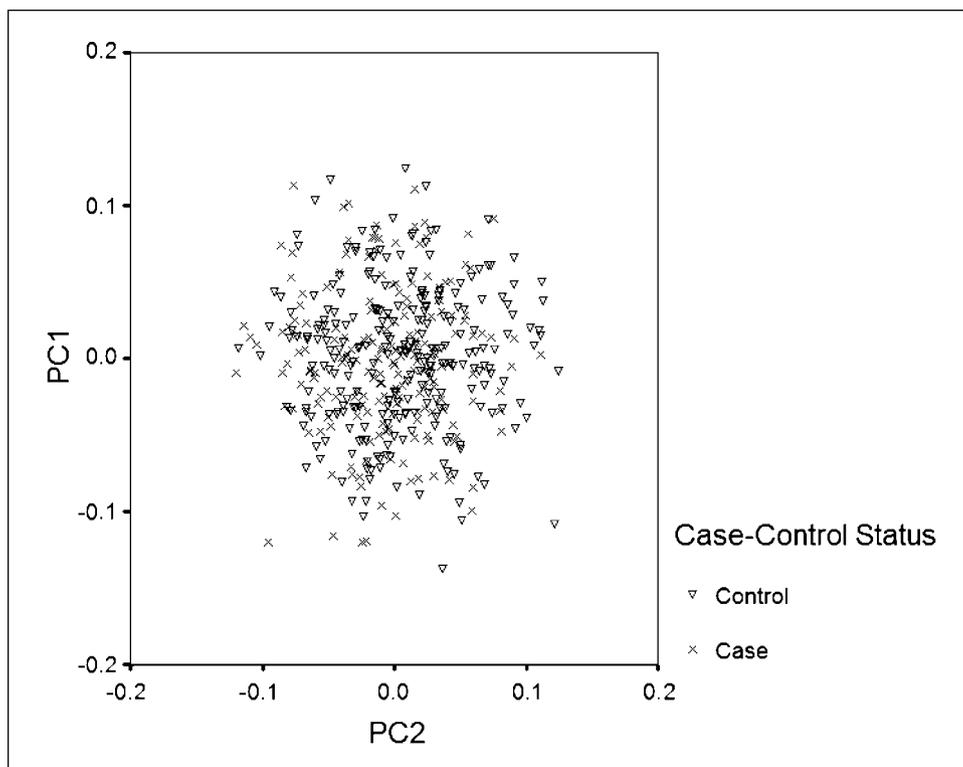


Figure 1. Eigenvectors of the first two principle components with the largest Eigenvalues plotted for the primary sample set using genotyping data of 152 SNPs.

Table 3. Genotyping results of the variants confirmed in the validation set

| Gene, variant | Primary set, <i>n</i> (%) | | | | Validation set, <i>n</i> (%) | | | | Joint set, <i>n</i> (%) | | | |
|------------------------------|---------------------------|-------------------|------|------------|------------------------------|---------------------|-----------------|-----------------------|-------------------------|---------------------|-----------------|-----------------------|
| | Cases (197) | Controls (254) | OR* | <i>P</i> * | Cases (549) | Controls (1,119) | OR [†] | <i>P</i> [‡] | Cases (746) | Controls (1,373) | OR [†] | <i>P</i> [‡] |
| ADH1B, Arg ⁴⁸ His | | | | | | | | | | | | |
| Arg/Arg | 129 (66.5) | 145 (57.3) | 1.00 | 1.00 | 361 (65.8) | 682 (61.0) | 1.00 | 1.00 | 490 (65.7) | 827 (60.2) | 1.00 | 1.00 |
| Arg/His | 59 (30.5) | 96 (37.9) | 0.66 | 0.04 | 173 (31.5) | 375 (33.5) | 0.91 | 1.00 | 232 (31.1) | 471 (34.3) | 0.84 | 0.10 |
| His/His | 6 (3.0) | 12 (4.8) | 0.70 | 0.17 | 15 (2.7) | 61 (5.5) | 0.59 | 0.0014 | 21 (2.8) | 73 (5.3) | 0.63 | 0.0002 |
| Dominant | | | 0.64 | 0.03 | | | 0.82 | 0.31 | | | 0.77 | 0.01 |
| Recessive | | | 0.61 | 0.33 | | | 0.35 | 0.0014 | | | 0.41 | 0.0004 |
| MGMT, rs7087131 | | | | | | | | | | | | |
| G/G | 137 (70.6) | 171 (68.7) | 1.00 | 1.00 | 417 (76.2) | 786 (70.3) | 1.00 | 1.00 | 554 (74.8) | 957 (70.0) | 1.00 | 1.00 |
| G/A | 56 (28.9) | 65 (26.1) | 1.05 | 0.82 | 115 (21.0) | 305 (27.3) | 0.74 | 0.07 | 171 (23.1) | 370 (27.1) | 0.81 | 0.06 |
| A/A | 1 (0.5) | 13 (5.2) | 0.30 | 0.002 | 15 (2.8) | 27 (2.4) | 0.96 | 1.00 | 16 (2.1) | 40 (2.9) | 0.78 | 0.09 |
| Dominant | | | 0.89 | 0.59 | | | 0.76 | 0.10 | | | 0.79 | 0.02 |
| Recessive | | | 0.10 | 0.002 | | | 0.98 | 1.00 | | | 0.63 | 0.13 |

**P* and OR were adjusted for age using multivariate logistic regression analysis.

[†]*P* and OR were adjusted for age and ethnicity using multivariate logistic regression analysis.

[‡]*P* is based on one-sided test and adjusted for age and ethnicity using multivariate logistic regression analysis and for multiple testing using Bonferroni correction.

The possibility that there is spurious population stratification in the Turkmens was addressed by principle component analysis using the Eigenstrat method (19) for identifying any population stratification in the primary sample set, which we had genotypes of 130 tagSNPs plus 22 functional variants for them. By plotting the Eigenvectors of the first two principle components with the largest Eigenvalues for the cases and the controls (Fig. 1), it appears that population stratification was not present. All analyses were done by SNP Variation Suit version 7 (Golden Helix).

Results

Initially, we genotyped 152 SNPs in a primary set of 197 cases and 254 controls. These were the first samples to be collected in the course of the project and all subjects were Turkmen. These SNPs included 22 functional variants of 15 genes and 130 tagSNPs from these genes. The results of the 17 functional variants and 3 slow-metabolizing alleles of NAT2 gene in the primary set of 451 samples are shown in Supplementary Table S1. The frequencies of the minor alleles of all variants, except one (ALDH2 Glu⁴⁸⁷Lys), were >1%. All were in Hardy-Weinberg equilibrium.

In addition, we selected and genotyped 130 tagSNPs of 15 different genes in the primary set of samples. These SNPs had not been studied previously in ESCC. Eight tagSNPs had minor allele frequencies of <0.01 and 3 tagSNPs were not in Hardy-Weinberg equilibrium. These 11 tagSNPs were excluded from the analysis. The remaining 119 tagSNPs were analyzed in the primary set of 197 cases and 254 controls.

Two functional variants showed associations with ESCC in the primary sample set with *P* values < 0.05 (the Arg⁴⁸His variant from ADH1B gene and c.870A>G from CCND1). Five tagSNPs, from five different genes, also showed significant associations in the primary set. These 7 SNPs (2 functional variants and 5 tagSNPs) were then genotyped in the validation set of samples. The results are reported separately for the primary set, the validation set, and the joint set (Table 3). Results were reported in the form of odds ratios (OR) for the three genotype classes, with the most common homozygous

genotype being the reference class. The cases and controls were different with regards to age and ethnic group (Table 1); therefore, all ORs were adjusted for age and ethnic group. Results were also reported assuming dominant and recessive models.

The strongest association was observed with the His allele of the Arg⁴⁸His variant of ADH1B gene. This allele was associated with a decreased risk of ESCC under a recessive model (Table 3). Compared with those who carried no His allele, those who carried two His alleles were at reduced risk (OR, 0.63; *P* = 0.0002; 95% confidence interval, 0.34-0.93).

This association with the rs7087131 variant of MGMT gene was borderline significant in the validation set after adjusting for multiple testing (Table 3). In the combined set, the A allele of the rs7087131 variant of MGMT gene was associated with a decreased risk of ESCC under a dominant model (OR, 0.79; 95% confidence interval, 0.64-0.96; *P* = 0.02). For the other 5 variants, the associations were failed to be confirmed in the validation set (Table 4).

Discussion

We studied 22 different functional variants of 15 genes, which were previously reported to be associated with ESCC. Notably, the ORs reported for these variants in the earlier studies were typically ≥ 2.0 (11). Given that the frequencies of the minor alleles of these variants in our population were almost all above 0.1 (Supplementary Table S1), the power of our study to detect ORs above 2.0 exceeded 90%. However, of the 22 variants tested, only two polymorphisms (ADH1B Arg⁴⁸His and CCND1 c.870A>G) showed significant associations at the *P* = 0.05 level in the primary sample set and ADH1B Arg⁴⁸His was then confirmed in the validation set.

In the joint sample set, the Arg⁴⁸His allele was protective under a recessive model (OR, 0.41; 95% confidence interval, 0.29-0.76; *P* = 4×10^{-4} ; Table 4). ADH1B is one of seven different genes that encode five different classes of alcohol dehydrogenase (ADH) enzymes. These enzymes are involved in the oxidation of alcohol

groups of a variety of substrates. ADH class I, which is responsible for the oxidation of ethanol to acetaldehyde, is a homodimer or heterodimer of three related subunits (α , β , and γ) encoded by ADH1A, ADH1B, and ADH1C, respectively (20). All ADH enzymes use NAD⁺/NADH as their oxidizing cofactor. The arginine residue at codon 48 of ADH1B is in close contact with its cofactor and seems to have a crucial role on modulating the ethanol oxidizing capability of the protein. When the arginine is replaced by histidine at this codon, the enzyme activity increases by 70- to 80-fold (20). In a rodent model, ADH1B, which has proline at this codon, is not capable of oxidizing ethanol (21).

In previous studies, the Arg allele of ADH1B consistently has been reported to increase the risk of ESCC by 1.6- to 4-fold among alcohol drinkers compared with individuals with two His alleles (10, 11, 22). The suggested mechanism is that alcohol and acetaldehyde persist for a longer time in the circulation of individuals with the Arg allele, who are slow metabolizers. These individuals tend to experience less flushing on exposure to alcohol and are prone to alcoholism (23). We confirmed that the presence of the Arg allele at codon 48 of the ADH1B protein increases the risk of ESCC by 2.4-fold in

comparison with two His alleles ($P = 4 \times 10^{-4}$). Considering that only 4.5% of our study subjects were alcohol drinkers, this finding suggests that in Iran the association is not due to metabolism of alcohol per se. Our observation is contrary to other studies that report no association between the Arg⁴⁸His variant of the ADH1B gene and the risk of ESCC among nondrinkers (22, 24, 25). Most studies that showed the association of Arg⁴⁸His variant of ADH1B with ESCC in alcohol drinkers also report an association between the Glu⁴⁸⁷Lys variant of ALDH2 and esophageal cancer (11). The ALDH2 protein with lysine residue at codon 487 is catalytically inactive and cannot metabolize acetaldehyde, a known carcinogen (23). We did not observe this association in Iran.

Our data suggest that, in Northern Iran, ESCC may be related to an environmental factor other than ethanol but is also metabolized by ADH1B. 1-Methylpyrene (1-MP) is a candidate for such an environmental factor. 1-MP is a carcinogenic polycyclic aromatic hydrocarbon (PAH), which is metabolized to 1-hydroxymethylpyrene (1-HMP) by cytochrome *P450* enzymes. 1-HMP could be activated by sulfotransferases to 1-sulfooxymethylpyrene (1-SMP), which is capable of forming

Table 4. Genotyping results of the five initially ESCC-associated variants that were not confirmed in the validation set

| Gene, variant | Primary set, <i>n</i> (%) | | | | Validation set, <i>n</i> (%) | | | |
|------------------|---------------------------|----------------|------|------------|------------------------------|------------------|-----------------|-----------------------|
| | Cases (197) | Controls (254) | OR* | <i>P</i> * | Cases (549) | Controls (1,119) | OR [†] | <i>P</i> [‡] |
| ALDH2, rs886205 | | | | | | | | |
| T/T | 99 (51.0) | 107 (43.2) | 1.00 | 1.00 | 289 (52.8) | 564 (50.4) | 1.00 | 1.00 |
| T/C | 80 (41.2) | 100 (40.3) | 0.87 | 0.52 | 209 (38.2) | 449 (40.1) | 0.92 | 1.00 |
| C/C | 15 (7.8) | 41 (16.5) | 0.62 | 0.004 | 49 (9.0) | 106 (9.5) | 0.97 | 1.00 |
| Dominant | | | 0.73 | 0.11 | | | 0.93 | 1.00 |
| Recessive | | | 0.42 | 0.005 | | | 1.00 | 1.00 |
| CCND1, c.870A>G | | | | | | | | |
| A/A | 53 (27.2) | 79 (31.6) | 1.00 | 1.00 | 158 (28.8) | 355 (31.8) | 1.00 | 1.00 |
| A/G | 83 (42.6) | 116 (46.4) | 1.09 | 0.72 | 276 (50.3) | 550 (49.2) | 1.02 | 1.00 |
| G/G | 59 (30.2) | 55 (22.0) | 1.30 | 0.04 | 115 (20.9) | 213 (19.0) | 1.04 | 1.00 |
| Dominant | | | 1.27 | 0.26 | | | 1.05 | 1.00 |
| Recessive | | | 1.61 | 0.03 | | | 1.09 | 1.00 |
| ECRG1, rs6552135 | | | | | | | | |
| A/A | 78 (40.2) | 69 (27.8) | 1.00 | 1.00 | 167 (31.2) | 339 (40.0) | 1.00 | 1.00 |
| A/G | 71 (36.6) | 117 (47.2) | 0.50 | 0.002 | 252 (47.1) | 524 (47.9) | 1.00 | 1.00 |
| G/G | 45 (23.2) | 62 (25.0) | 0.80 | 0.09 | 116 (21.7) | 231 (21.1) | 0.99 | 1.00 |
| Dominant | | | 0.55 | 0.004 | | | 1.00 | 1.00 |
| Recessive | | | 0.94 | 0.80 | | | 0.98 | 1.00 |
| ERCC2, rs3916874 | | | | | | | | |
| G/G | 143 (73.0) | 152 (60.8) | 1.00 | 1.00 | 379 (69.2) | 739 (66.4) | 1.00 | 1.00 |
| G/C | 47 (24.0) | 87 (34.8) | 0.56 | 0.008 | 152 (27.7) | 337 (30.3) | 0.93 | 1.00 |
| C/C | 6 (3.0) | 11 (4.4) | 0.80 | 0.39 | 17 (3.1) | 37 (3.3) | 0.92 | 1.00 |
| Dominant | | | 0.57 | 0.007 | | | 0.92 | 1.00 |
| Recessive | | | 0.76 | 0.60 | | | 0.85 | 1.00 |
| TP53, rs1625895 | | | | | | | | |
| G/G | 112 (57.4) | 167 (66.8) | 1.00 | 1.00 | 318 (58.1) | 689 (61.6) | 1.00 | 1.00 |
| G/A | 73 (37.4) | 75 (30.0) | 1.53 | 0.04 | 197 (36.0) | 365 (32.6) | 1.07 | 1.00 |
| A/A | 10 (5.1) | 8 (3.2) | 1.35 | 0.23 | 32 (5.9) | 65 (5.8) | 1.00 | 1.00 |
| Dominant | | | 1.55 | 0.03 | | | 1.07 | 1.00 |
| Recessive | | | 1.57 | 0.36 | | | 0.99 | 1.00 |

**P* and OR were adjusted for age using multivariate logistic regression analysis.

†OR was adjusted for age and ethnicity using multivariate logistic regression analysis.

‡*P* is based on one-sided test and adjusted for age and ethnicity using multivariate logistic regression analysis and for multiple testing using Bonferroni correction.

DNA adducts, but most of 1-HMP is inactivated by ADH1B and excreted in urine and does not form DNA adducts (26, 27).

PAHs have been shown to be esophageal carcinogens in both animal and human studies. In a 2-year feeding study in mice, a dose-response relationship was shown between benzo[*a*]pyrene intake and ESCC incidence (28). Occupational exposure to PAHs in Sweden was reported to double the risk of ESCC (29). Other lines of evidence that support the contribution of PAHs in the pathogenesis of ESCC come from a high-risk region in China and include high urinary levels of PAH metabolites in residents (30), increased concentrations of PAHs in cooked and uncooked foods in the region (28), and the presence of PAH-DNA adducts in ESCC tissue samples of patients (31).

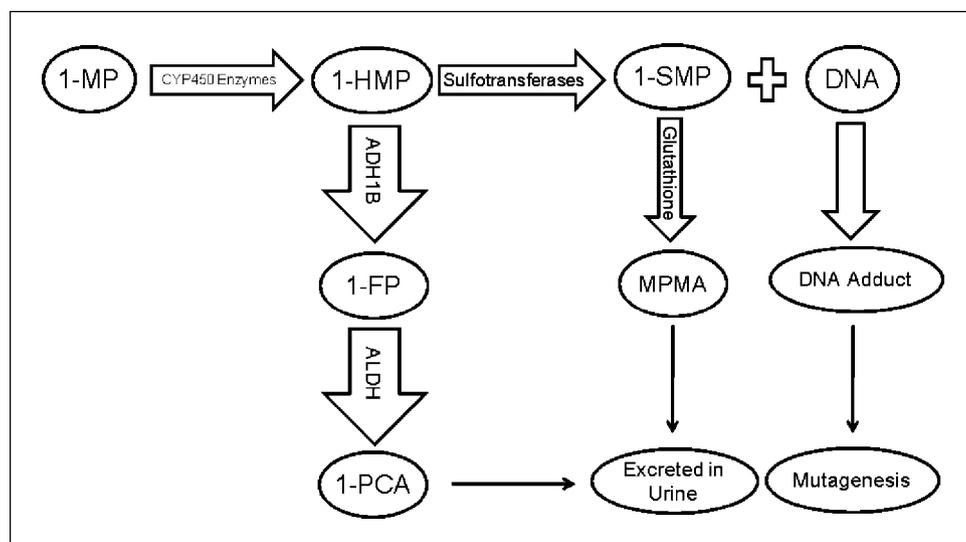
PAHs require metabolic activation to become carcinogens. Purely aromatic PAHs, such as benzo[*a*]pyrene, are activated via the diol-epoxide pathway (32), but alkylated PAHs, such as 1-MP, are metabolized by an alternative pathway named the aralkylating hydrocarbon pathway (33). In this pathway (Fig. 2), 1-MP is hydroxylated to 1-HMP by cytochrome *P*450 enzymes such as CYP1A1 (34). 1-HMP subsequently forms reactive sulfuric acid ester (1-SMP) by the action of sulfotransferases (35, 36). 1-SMP forms DNA adducts and has been shown to be mutagenic in rats (33) and in *Salmonella typhimurium* (37). 1-SMP is also reacts with glutathione and is excreted in urine as methylpyrenyl mercapturic acid (MPMA). The urine MPMA constitutes only 2% of the administered 1-HMP to rats; 80% of it is excreted as free 1-pyrenylcarboxylic acid (1-PCA) and its glucuronic acid or other conjugates (26). 1-PCA is an inactivate metabolite of 1-HMP, which has been oxidized by ADH1B (27). Therefore, ADH1B competes with sulfotransferases for the metabolism of 1-HMP and ADH1B activity decreases the formation of DNA adducts. If ADH1B is the major enzyme for the inactivation of 1-HMP, then other substrates of ADH1B are expected to interfere with inactivation of 1-HMP and increase the amount of its activated metabolite. This has been exhibited by concurrent administration of 1-HMP and ethanol to rats, thereby enhancing the hepatic level of PAH-DNA adducts by 15-fold (26). The competition of ethanol with 1-HMP for oxidation by ADH1B could also explain the synergistic carcinogenic effect of alcohol and cigarette smoke (which is rich in 1-MP, the precursor of 1-HMP; ref. 27). It is possible that individuals with two slow-metabolizing Arg alleles at codon 48 of ADH1B gene do not

inactivate 1-HMP as efficiently as those who carry one or more His alleles at this codon. In this case, more 1-HMP would be available for conversion to 1-SMP by sulfotransferases; consequently, more DNA adducts would be formed (Fig. 2).

The rs7087131 variant of MGMT was also found to be associated with ESCC under a dominant model. However, this association was of borderline significance. *O*⁶-methylguanine is formed by alkylating agents such as nitrosamines, which are associated with an increased risk of ESCC (38). *O*⁶-methylguanine has been found in the epithelial cells of the esophagus of individuals living in the high-risk region of China (39). Unrepaired *O*⁶-methylguanine tends to pair with thymine and introduces G:C to A:T transitions during cell replication (40). *O*⁶-methylguanine-DNA methyltransferase (MGMT) repairs *O*⁶-methylguanine by transferring the methyl group to a cysteine residue at codon 145 of the enzyme (41). Binding of methyl group to the cytosine residues of MGMT is an irreversible reaction, and after repairing each single *O*⁶-methylguanine, the protein is ubiquitinated and degraded and needs to be replaced by a new protein. Therefore, it is expected that cells with low expression of MGMT cannot repair *O*⁶-methylguanine efficiently. Low expression of MGMT in cells can result from either promoter hypermethylation of the gene or by *cis*-acting nucleotide polymorphisms, which change the binding sites of key regulatory factors (42). A group of SNPs at the 5' half of the gene were described to be in association with MGMT activity in WBC (42). The rs7087131 variant is also located at the 5' half of the MGMT gene in the intronic region between exons 2 and 3 and directly or indirectly might affect expression level of the gene and contribute to the risk of ESCC in Iran. In our study, the substitution of G by A nucleotide in this variant decreased the risk of ESCC under a dominant model (OR, 0.79; 95% confidence interval, 0.64-0.96; *P* = 0.02).

There are several limitations to our study. The sample size was relatively small and we did not have sufficient power to detect associations of a modest size, for example, ORs of <1.5. The primary data set was composed of Turkmens, and if an association were to be present only in non-Turkmens, it would not be discovered using this strategy. Similarly, if an association were present only in Turkmens, it would fail to be confirmed in the validation set. We have collected data on family history for most of the cases, but because of the high fatality rate of esophageal cancer, a DNA sample was available for the proband only.

Figure 2. Suggested metabolic pathway for 1-MP. 1-MP is first metabolized to 1-HMP by cytochrome *P*450 (*CYP450*) enzymes. The majority (~80%) of 1-HMP is detoxified by alcohol dehydrogenase (*ADH1B*) to 1-formylpyrene (1-FP) and then to 1-PCA by aldehyde dehydrogenase (*ALDH*). 1-PCA and its glucuronic acid conjugates are excreted in urine. Some of (~20%) 1-HMP are transformed to 1-SMP of which the majority form DNA adducts and some of it (~2%) are detoxified by nucleophile glutathione and form MPMA.



In conclusion, this study extends our earlier observations that genetic factors play an important role in the high incidence of esophageal cancer in the northeast of Iran. A strong and significant association was seen with the Arg⁴⁸His allele of ADH1B gene among both Turkmens and non-Turkmens. In several previous studies, this variant was associated with the risk of ESCC, primarily among alcohol drinkers (10, 11, 22). Interestingly, in this study, the variant is related to the risk of ESCC in alcohol nondrinkers, suggesting that other carcinogens might be causative.

It is important that this observation be confirmed in larger series of patients. To date, we have strong evidence that both BRCA2 and ADH1B contribute to the burden of esophageal cancer in Northern Iran, but the known variants in these two genes are not sufficient to explain the high incidence of esophageal cancer in the region, or the high familial relative risk, and further studies are needed.

References

- Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
- Kmet J, Mahboubi E. Esophageal cancer in the Caspian littoral of Iran: initial studies. *Science* 1972;175:846–53.
- Saidi F, Sepehr A, Fahimi S, et al. Oesophageal cancer among the Turkmans of northeast Iran. *Br J Cancer* 2000;83:1249–54.
- Semmani S, Sadjadi A, Fahimi S, et al. Declining incidence of esophageal cancer in the Turkmen Plain, eastern part of the Caspian littoral of Iran: a retrospective cancer surveillance. *Cancer Detect Prev* 2006;30:14–9.
- Islami F, Kamangar F, Aghcheli K, et al. Epidemiologic features of upper gastrointestinal tract cancers in northeastern Iran. *Br J Cancer* 2004;90:1402–6.
- Akbari MR, Malekzadeh R, Nasrollahzadeh D, et al. Familial risks of esophageal cancer among the Turkmen population of the Caspian littoral of Iran. *Int J Cancer* 2006;119:1047–51.
- Akbari MR, Malekzadeh R, Nasrollahzadeh D, et al. Germline BRCA2 mutations and the risk of esophageal squamous cell carcinoma. *Oncogene* 2008;27:1290–6.
- Sumer F, Oguzlar (Turkmenler), 5th ed. Istanbul: Turk Dunyasi Arastirmalari Yakfi; 1999.
- Ke L. Mortality and incidence trends from esophageal cancer in selected geographic areas of China CIRCA 1970–90. *Int J Cancer* 2002;102:271–4.
- Xing D, Tan W, Lin D. Genetic polymorphisms and susceptibility to esophageal cancer among Chinese population [review]. *Oncol Rep* 2003;10:1615–23.
- Hiyama T, Yoshihara M, Tanaka S, Chayama K. Genetic polymorphisms and esophageal cancer risk. *Int J Cancer* 2007;121:1643–58.
- Pourshams A, Khademi H, Malekshah AF, et al. Cohort profile: the Golestan Cohort Study—a prospective study of oesophageal cancer in northern Iran. *Int J Epidemiol. Epub* 2009 Mar 30.
- Sadjadi A, Malekzadeh R, Derakhshan MH, et al. Cancer occurrence in Ardabil: results of a population-based cancer registry from Iran. *Int J Cancer* 2003;107:113–8.
- Bugni JM, Han J, Tsai MS, Hunter DJ, Samson LD. Genetic association and functional studies of major polymorphic variants of MGMT. *DNA Repair (Amst)* 2007;6:1116–26.
- Morita S, Yano M, Tsujinaka T, et al. Association between genetic polymorphisms of glutathione *S*-transferase P1 and *N*-acetyltransferase 2 and susceptibility to squamous-cell carcinoma of the esophagus. *Int J Cancer* 1998;79:517–20.
- International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–320.
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
- Fallin D, Schork NJ. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet* 2000;67:947–59.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
- Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health* 2007;30:5–13.
- Hög JO, Hedberg JJ, Strömberg P, Svensson S. Mammalian alcohol dehydrogenase—functional and structural implications. *J Biomed Sci* 2001;8:71–6.
- Hashibe M, McKay JD, Curado MP, et al. Multiple ADH genes are associated with upper aerodigestive cancers. *Nat Genet* 2008;40:707–9.
- Yokoyama A, Omori T. Genetic polymorphisms of alcohol and aldehyde dehydrogenases and risk for esophageal and head and neck cancers. *Jpn J Clin Oncol* 2003;33:111–21.
- Yang SJ, Wang HY, Li XQ, et al. Genetic polymorphisms of ADH2 and ALDH2 association with esophageal cancer risk in southwest China. *World J Gastroenterol* 2007;13:5760–4.
- Lee CH, Lee JM, Wu DC, et al. Carcinogenic impact of ADH1B and ALDH2 genes on squamous cell carcinoma risk of the esophagus with regard to the consumption of alcohol, tobacco and betel quid. *Int J Cancer* 2008;122:1347–56.
- Ma L, Kuhlow A, Glatt H. Ethanol enhances the activation of 1-hydroxymethylpyrene to DNA adduct-forming species in the rat. *Polycyclic Aromatic Compounds* 2002;22:933–46.
- Kollock R, Meinel W, Schneider H, et al. Efficient oxidation of promutagenic hydroxymethylpyrenes by cDNA-expressed human alcohol dehydrogenase ADH2 and its inhibition by various agents. *Biochem Pharmacol* 2008;75:527–37.
- Roth MJ, Guo-King W, Lewin KJ, et al. High levels of carcinogenic polycyclic aromatic hydrocarbons present within food from Linxian, China may contribute to that region's high incidence of oesophageal cancer. *Eur J Cancer* 1998;34:757–8.
- Gustavsson P, Jakobsson R, Johansson H, Lewin F, Norell S, Rutkvist LE. Occupational exposures and squamous cell carcinoma of the oral cavity, pharynx, larynx, and oesophagus: a case-control study in Sweden. *Occup Environ Med* 1998;55:393–400.
- Roth MJ, Qiao Y, Rothman N, et al. High urine 1-hydroxypyrene glucuronide concentrations in Linxian, China, an area of high risk for squamous oesophageal cancer. *Biomarkers* 2001;6:381–6.
- Van-Gijssel HE, Divi RL, Olivero OA, et al. Semi-quantitation of polycyclic aromatic hydrocarbon-DNA adducts in human esophagus by immunohistochemistry and the automated cellular imaging system. *Cancer Epidemiol Biomarkers Prev* 2002;11:1622–9.
- Thakker DR, Yagi H, Levin W, Wood AW, Conney AH, Jerina DM. Polycyclic aromatic hydrocarbons: metabolic activation to ultimate carcinogens. In: Anders MW, editor. *Bioactivation of foreign compounds*. Academic Press; 1985. p. 177–242.
- Horn J, Flesher JW, Lehner AF. 1-Sulfooxymethylpyrene is an electrophilic mutagen and ultimate carcinogen of 1-methyl- and 1-hydroxymethylpyrene. *Biochem Biophys Res Commun* 1996;228:105–9.
- Engst W, Landsiedel R, Hermersdörfer H, Doehmer J, Glatt H. Benzyl hydroxylation of 1-methylpyrene and 1-ethylpyrene by human and rat cytochromes *P450* individually expressed in V79 Chinese hamster cells. *Carcinogenesis* 1999;20:1777–85.
- Surh Y-J, Miller JA. Roles of electrophilic sulfuric acid ester metabolites in mutagenesis and carcinogenesis by some polynuclear aromatic hydrocarbons. *Chem Biol Interact* 1994;92:351–62.
- Glatt HR, Meinel W, Kuhlow A, Ma L. Metabolic formation, distribution and toxicological effects of reactive sulphuric acid esters. *Nova Acta Leopoldina* 2003;329:151–61.
- Glatt HR, Seidel A, Harvey RG, Coughtrie MWH. Activation of benzylic alcohols to mutagens by human hepatic sulphotransferases. *Mutagenesis* 1994;9:553–7.
- Lu SH, Chui SX, Yang WX, Hu XN, Guo LP, Li FM. Relevance of *N*-nitrosamines to oesophageal cancer in China. *IARC Sci Publ* 1991;105:11–7.
- Umbenhauer D, Wild CP, Montesano R, et al. O(6)-methyldeoxyguanosine in oesophageal DNA among individuals at high risk of oesophageal cancer. *Int J Cancer* 1985;36:661–5.
- Snow ET, Foote RS, Mitra S. Base-pairing properties of *O*⁶-methylguanine in template DNA during *in vitro* DNA replication. *J Biol Chem* 1984;259:8095–100.
- Olsson M, Lindahl T. Repair of alkylated DNA in *Escherichia coli*. Methyl group transfer from *O*⁶-methylguanine to a protein cysteine residue. *J Biol Chem* 1980;255:10569–71.
- Margison GP, Heighway J, Pearson S, et al. Quantitative trait locus analysis reveals two intragenic sites that influence *O*⁶-alkylguanine-DNA alkyltransferase activity in peripheral blood mononuclear cells. *Carcinogenesis* 2005;26:1473–80.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 4/4/09; revised 6/16/09; accepted 7/2/09; published OnlineFirst 10/13/09.

Grant support: Canadian Institutes of Health Research and Digestive Disease Research Center of the Tehran University of Medical Sciences.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Dr. Jafar Bashiri [case-control study center (Aras Clinic) in Ardabil City], Safura Kor [case-control study center (Atrak Clinic)], Goharshad Gogiani and Mina Bahrami (cohort study center in Gonbad City) for help with study subject recruitment, Cheryl Crozier (Analytical Genetics Technology Centre, University of Toronto) for the kind help, and Alexandre Belisle (Genotyping Platform of the McGill University and Genome Quebec Innovation Centre) for help in performing the genetic tests of this study.