Allelic Imbalance at rs6983267 Suggests Selection of the Risk Allele in Somatic Colorectal Tumor Evolution

Sari Tuupanen,1 Iina Niittymäki,1 Kari Nousiainen,2 Sakari Vanharanta,1 Jukka-Pekka Mecklin,4 Kyösti Nuorva,1 Heikki Järvinen,3 Sampsu Hautaniemi,3 Auli Karhu,1 and Lauri A. Aaltonen1

1Department of Medical Genetics, Genome-Scale Biology Research Program, Biomedicum Helsinki; 2Computational Systems Biology Laboratory, Genome-Scale Biology Research Program, Institute of Biomedicine, University of Helsinki; and
3The Second Department of Surgery, Helsinki University Hospital, Helsinki, Finland; and Departments of
4Surgery and Pathology, Jyväskylä Central Hospital, Jyväskylä, Finland

Abstract
A common single nucleotide polymorphism (SNP), rs6983267, at 8q24.21 has recently been shown to associate with colorectal cancer (CRC). Three independent SNP association studies showed that rs6983267 contributes to CRC with odds ratios (OR) of 1.17 to 1.22. Here, we genotyped a population-based series of 1,042 patients with CRC and 1,012 healthy controls for rs6983267 and determined the contribution of SNP to CRC in Finland, using germ line DNA, as well as the respective cancer DNA in heterozygous patients. The comprehensive clinical data available from the 1,042 patients and their first-degree relatives enabled us to thoroughly examine the possible association of this variant with different clinical features. As expected, a significant association between the G allele of rs6983267 and CRC [OR, 1.22; 95% confidence interval (CI), 1.08–1.38; P = 0.0018] was found, confirming the previous observations. A trend towards association of the G allele with microsatellite-stable cancer (OR, 1.37; 95% CI, 1.02–1.85; P = 0.04) and family history of cancers other than CRC was seen (OR, 1.20; 95% CI, 1–1.43; P = 0.05). Four hundred and sixty-six GT heterozygotes identified in this study were analyzed for allelic imbalance at rs6983267 in the respective cancer DNA. One hundred and one tumors showed allelic imbalance (22%). The risk allele G was favored in 67 versus 34 tumors (P = 0.0007). This finding implicates that the underlying germ line genetic defect in 8q24.21 is a target in the somatic evolution of CRC. [Cancer Res 2008;68(1):14–7]

Introduction
Inherited susceptibility is likely to play a role in 35% of all colorectal cancers (CRC; ref. 1). Approximately 5% of CRCs are caused by mutations in high-penetrance predisposition genes such as APC, MLH1, MSH2, and MYH (2). Common low-penetrance alleles could explain much of the remaining inherited predisposition.

Recently, technological advances in single nucleotide polymorphism (SNP) genotyping have allowed large-scale case-control association studies in search of low-penetrance alleles. SNP rs6983267 at 8q24.21 has been identified as a new common susceptibility variant for CRC. Tomlinson et al. (3) performed an association study with 550,000 tag SNPs and a case-control sample series from United Kingdom, and identified the association of rs6983267 with CRC. Pooled data from 7,954 CRC cases and 6,206 controls indicated that rs6983267 was associated with CRC with an odds ratio (OR) of 1.21 [95% confidence interval (CI), 1.15–1.27; P = 1.27 × 10−11]. In an independent study, Haiman et al. (4) evaluated prostate cancer–associated SNPs at 8q24 in CRC predisposition. By analyzing 1,807 affected individuals and 5,511 controls that were derived from five populations, they found that rs6983267 was also a risk variant for CRC (OR, 1.22; 95% CI, 1.12–1.32; P = 4.4 × 10−6). Moreover, a multistage association study comprising 7,480 CRC cases and 7,779 controls from the United States, Canada, and Europe pinpointed chromosome 8q24 as a susceptibility locus for CRC (5). The association was obtained with rs10505477 (OR, 1.17; 95% CI, 1.12–1.23; P = 3.16 × 10−7) that is in tight linkage disequilibrium with rs6983267. Association of rs10505477 with CRC was found also in a population-based case-control study from Israel (6).

The frequency of the risk allele at rs6983267 seemed to be high among all populations, varying from 31% in Native Hawaiians to 85% in African Americans and being 50% in Caucasians (4). Some differences in CRC risks between populations exist. Native Hawaiians and African Americans seem to have the highest risks, with ORs of 1.59 and 1.37, respectively. In populations of European origin, ORs in the United Kingdom and in European Americans were 1.21 and 1.28, respectively (3, 4).

The location of rs6983267 in the intergenic region at 8q24.21 sets challenges for establishing the mechanism by which this locus promotes CRC tumorigenesis. To our knowledge, no previous studies have been published on the possible role of this germ line predisposition SNP in somatic tumor evolution. In this study, we have examined the contribution of rs6983267 to CRC in Finland and genotyped a population-based series of 1,042 Finnish CRC cases and 1,012 healthy population-matched controls for rs6983267. The comprehensive clinical data available from the patients with CRC and their first-degree relatives was used in studying the possible association of the SNP with different clinical features. To study whether the predisposition allele G is favored during somatic CRC development, implicating the amplification of the mutant allele or loss of the wild-type allele, allelic imbalance (AI) analysis was conducted using tumor DNA from patients heterozygous for the SNP.

Materials and Methods

Study subjects. The study material consisted of population-based series of 1,042 Finnish CRC patients. This sample series, including both normal and tumor tissue specimens, was collected from 1994 to 1998, and is described in detail in previous reports (7, 8). The source of tumor DNAs was fresh-frozen tissue. Patient information and samples were obtained with

Requests for reprints: Lauri A. Aaltonen, Department of Medical Genetics, Biomedicum Helsinki, University of Helsinki, P.O. Box 63, FIN-00014 Helsinki, Finland. Phone: 358-91912-5595; Fax: 358-91912-5105; E-mail: lauri.aaltonen@helsinki.fi.

doi:10.1158/0008-5472.CAN-07-5766
informed consent and Ethical Review Board approval. The 1,012 controls used in this study were anonymous Finnish cancer-free blood donors obtained from the Finnish Red Cross Blood Transfusion Service.

Genotyping. Genotyping of CRC cases and healthy controls for rs6983267 was carried out by genomic sequencing. Genomic DNA was amplified with PCR using primers 5′-CTGACCCCTGCTAATTTGCT-3′ and 5′-CAGTCTAAGGCCCTTACCT-3′. PCR products were sequenced directly using Applied Biosystems BigDye v3.1 sequencing chemistry and ABI3730 Automatic DNA sequencer.

Analysis of AI. To assess AI, tumor DNA from individuals heterozygous for rs6983267 (n = 478) was sequenced. All tumor samples were histologically verified and 63% of the samples displayed 70% or more carcinoma tissue. AI was scored by comparing the ratios of the allele peak heights in sequencing graphs between normal and tumor samples as described previously (9, 10). The cutoffs for AI were <0.60 and >1.67 (10).

Statistical methods. Statistical analyses were performed using R software, version 2.5.1.2 Differences between allele frequencies in CRC cases and controls, as well as association of rs6983267 with clinical and tumor characteristics were undertaken with χ2 test with 1 df. Associations with clinical features were done with case-only analyses by dividing the patients into two groups according to sex, age (median age of diagnosis, 69 years, as separator), microsatellite instability (MSI) status [microsatellite unstable (MSU) versus MSI], family history of CRC, family history of other cancer, Duke’s stage (A–B versus C–D), and site (colon versus rectum).

Results

A total of 996 patients with CRC and 1,012 population-matched healthy controls were successfully genotyped for rs6983267. The frequency of G allele among controls was 52.5%, similar to previously published frequencies in Caucasian populations (3, 4). A significant difference in allele frequencies between individuals with CRC and healthy controls was found (OR, 1.22; 95% CI, 1.08–1.38; P = 0.0018; Table 1), indicating that also in Finns, the G allele of rs6983267 is associated with CRC. The OR for GG homozygous patients was 1.48 (95% CI, 1.15–1.89; P = 0.0022) and for GT heterozygotes it was 1.21 (95% CI, 0.96–1.52; P = 0.11).

The detailed patient data available enabled us to study the possible association of rs6983267 with clinical characteristics. The risk allele rs6983267 G was found to be somewhat more frequent in patients with MSS rather than MSI CRC (P = 0.04; Table 1). The G allele was also slightly more common in individuals with one or more first-degree relatives diagnosed with cancers other than CRC (P = 0.05). No evidence for association of rs6983267 G with gender (P = 0.73), familial CRC (P = 0.52), age at diagnosis (P = 0.64), Duke’s stage (P = 0.14), or site of the tumor (P = 0.26) was obtained (Table 1). Subsequently, we tested for possible differences in the clinical features between GG homozygous and GT heterozygous patients, and similar results were obtained from this analysis: GG homozygotes were associated with MSS CRC (MSS versus MSI: P = 0.03; OR, 1.88; 95% CI, 1.06–3.33) and with the occurrence of malignancies other than CRC in the family (P = 0.04; OR, 1.46; 95% CI, 1.01–2.09). The P values have not been corrected for multiple testing.

Tumor tissue DNA was available in 473 out of 478 heterozygous CRC patients identified in this study, of which 466 (97%) were successfully analyzed. 101 out of 466 (22%) tumors showed AI (Fig. 1). Relative reduction of G allele was observed in 34 out of 101 (34%) tumors and reduction of T allele in 67 out of 101 (66%).

---

Table 1. Association of rs6983267 with CRC risk and clinical characteristics

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>χ²</th>
<th>P</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>232</td>
<td>0.525</td>
<td>0.475</td>
<td></td>
</tr>
<tr>
<td>GT</td>
<td>283</td>
<td>0.574</td>
<td>0.426</td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>329</td>
<td>0.578</td>
<td>0.422</td>
<td></td>
</tr>
</tbody>
</table>

*Group versus healthy controls.
However, as expected, the frequency of AI was higher in patients than in the rest of the samples, no significant associations were found. The content of malignant tissue in the sample was 80%. 

**Discussion**

The G allele of rs6983267 is associated with elevated risk of CRC also in Finns. GG homozygous cases had a higher risk than patients heterozygous for the risk allele, with ORs being 1.48 and 1.21, respectively. Almost identical ORs have been obtained from the United Kingdom: 1.47 (95% CI, 1.34–1.62) for GG homozygotes and 1.27 (95% CI, 1.16–1.39) for heterozygotes (3).

We tested the association of rs6983267 with clinical and tumor characteristics and found a trend towards the association of the SNP with MSS CRC (unadjusted $P = 0.04$). In contrast, in the study by Tomlinson et al. (3), no association between rs6983267 and MSI status was found ($P = 0.55$), but detailed data, e.g., number of subjects, was not presented. Also, a trend towards a higher frequency of the SNP in patients with a family history of malignancies other than CRC was observed (unadjusted $P = 0.05$). No significant evidence for association of the variant with any other clinical characteristics in our study or in the previous studies was obtained (3, 4, 6).

The observation that rs6983267 could increase the risk of other malignancies in the family prompted us to study the cancer spectrum in the families of GG homozygous CRC patients compared with TT homozygotes. The proportions of the most frequent cancers were (GG versus TT) 14.6% versus 14% for basalaioma, 13.6% versus 11% for lung cancer, 10.8% versus 6.6% for gastric cancer, 10.5% versus 9.6% for breast cancer, and 6.2% versus 7.2% for prostate cancer. This indicates no striking site-specific differences. Previously, rs6983267 has been shown to associate with prostate cancer (12) but not with breast cancer (13) or chronic lymphocytic leukemia (14).

To examine rs6983267 in somatic CRC evolution, we conducted an AI study. The amplification of chromosomal region 8q24 is a consistent finding in colorectal tumors and cell lines, both in MSI and MSS cancers (15, 16). Chromosomal imbalances are generally more frequent in MSS than in MSI cancers and the target of amplification in 8q24 seems to differ; 8q24.21 being the target in MSS cancers and 8q24.3 in MSI cancers (15). In our AI study, the use of fresh-frozen— and not paraffin-embedded—tumor material provided robust data. AI was detected in 22% of the tumors (101 of 466) and was associated with MSS CRC ($P = 0.013$). Risk allele G was more frequently favored ($P = 0.0007$), suggesting preferential amplification of the $G$ allele or loss of the $T$ allele, during tumorigenesis. Because the 8q24 region is frequently amplified, $G$ allele amplification seems more likely. Based on the high frequency of the $G$ allele in the population, it is well possible that all $G$ allele carriers do not carry the true predisposing defect. Somatic amplification of the $G$ allele could be a sign of true germ line predisposition, and serves as a useful tag to enrich patients with the causative germ line mutation.

The observation that the frequencies of both $G$ allele and AI were higher in MSS cancers compared with MSI cancers could suggest a less important role of this locus in MSI tumorigenesis. If correct, this could be due to the demand of gross somatic genetic changes (loss or amplification) at the susceptibility locus; such events are much rarer in MSI CRCs (11).

SNP rs6983267 is distant from the coding sequences, in the middle of a 19-kb haplotype block with high linkage disequilibrium. The 8q24.21 genomic region is gene-poor. The nearest gene, the pseudogene POU5F1P1, resides 14,807 bp distally. The well-known oncogene MYC, 335 kb from the rs6983267, follows POU5F1P1. The nearest proximal gene is FAM84B that lies 849 kb from the SNP. Because rs6983267 does not reside in the coding sequence, the mechanism by which it affects CRC tumorigenesis is challenging to establish. It is well possible that the real causal variant is in linkage disequilibrium with rs6983267 and is as yet unidentified. The underlying genomic sequence can also comprise a cis- or trans-regulatory region and the SNP could therefore have effects on gene expression. Cis-acting variation can involve promoters or enhancers that lie from a few to hundreds of kilobases upstream of the gene (17). Identification of target genes for such variations is possible by studying the correlation of the genotype and expression of nearby genes, whereas effects of trans-acting variants are more difficult to discover. Recently, regulatory SNPs in nongenic regions underlying complex human diseases have been identified (18, 19). The well-established role of MYC in colorectal neoplasia makes it the most promising candidate gene at 8q24.21. Because MYC lies outside the linkage disequilibrium block, rs6983267 could be associated with MYC with some regulatory effect.

In summary, we have further confirmed the association of rs6983267 with increased risk of CRC. The observation that AI at rs6983267 favors the risk allele, suggests that this locus is a target for additional somatic changes in CRCs. Further characterization

![Figure 1. An example of a heterozygous CRC case with AI favoring the G allele. Allele ratio, obtained from sequence graphs of normal (A) and tumor tissue (B) was 2.36, representing the average score among cases with T allele reduction. The content of malignant tissue in the sample was 80%.](image-url)
of the 8q24.21 region targeted by Al could reveal the underlying effect by which the rs6983267 locus is involved in the genesis of CRC.

Acknowledgments
Received 10/5/2007; revised 11/6/2007; accepted 11/7/2007.

Grantsupport: Academy of Finland (Finnish Center of Excellence Program 2006-2011), the Finnish Cancer Society, the Sigrid Juselius Foundation, the European Commission (LSHG-CT-2004-512142), Ida Montin Foundation (S. Tuupanen), and Biomedicum Helsinki Foundation (S. Tuupanen).

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 Sini Marttinen, Kirsi Pyhäläinen, and Taula Lehtinen for collecting the clinical data and Mairi Kuris, Mikko Aho, and Iina Vuoristo for technical assistance.

References
Allelic Imbalance at rs6983267 Suggests Selection of the Risk Allele in Somatic Colorectal Tumor Evolution

Sari Tuupanen, Iina Niittymäki, Kari Nousiainen, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/68/1/14

Cited articles
This article cites 18 articles, 8 of which you can access for free at:
http://cancerres.aacrjournals.org/content/68/1/14.full.html#ref-list-1

Citing articles
This article has been cited by 18 HighWire-hosted articles. Access the articles at:
/content/68/1/14.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.