**The E-cadherin −347G→GA promoter polymorphism and its effect on transcriptional regulation**

Yong Shin1±2, Il-Jin Kim1±, Hio Chung Kang1, Jae-Hyun Park1, Hye-Rin Park1, Hyo-Won Park1, Mi Ae Park2, Jong Soo Lee2, Kyong-Ah Yoon1, Ja-Lok Ku1 and Jae-Gahb Park1±3

1Korean Hereditary Tumor Registry, Cancer Research Center and Cancer Research Institute, Seoul National University and 2Research Institute and Hospital, National Cancer Center, Goyang, Gyeonggi, South Korea
3To whom correspondence should be addressed at: National Cancer Center, 809 Madu-dong, Ilsan-gu, Goyang, Gyeonggi, South Korea

E-cadherin plays a critical role in epithelial cell–cell adhesion and maintenance of tissue architecture. Loss of E-cadherin expression in humans has been associated with cancer, and a number of cancer-related mutations have been identified. Here, we sought to investigate whether the −347G→GA single nucleotide polymorphism affects the transcriptional activity of the E-cadherin gene. First, we measured the promoter activity of the −347G→GA polymorphism using a dual luciferase reporter assay and electrophoretic mobility shift assay (EMSA). The dual luciferase reporter assay showed that the GA allele decreased the transcriptional efficiency by 10-fold ($P < 0.001$) compared with the G allele. Similarly, EMSA revealed that the GA allele had a weak transcription factor binding strength compared with the G allele. We then examined the frequency of this polymorphism in familial gastric cancer (FGC) patients by denaturing high-performance liquid chromatography. We found that the E-cadherin genotype (−347G/GA heterozygous or GA homozygous) was associated with FGC patients ($P < 0.05$) compared with the G homozygous genotype. Taken together, these results suggest that the GA allele may cause weak transcription factor binding affinity and low transcriptional activity in E-cadherin expression.

**Introduction**

Gastric cancer is one of the most common cancers worldwide. Although the occurrence rate of gastric cancer has decreased in recent years, the incidence of the disease is still high in Asian countries such as Korea and Japan (1). However, relatively little is known regarding genetic susceptibility in the pathogenesis of gastric cancer (2). Mutations in the calcium-dependent cell adhesion molecule, E-cadherin, have been associated with the early development of gastric cancer (3). E-cadherin germ-line mutations were first identified in New Zealand Maori families with early-onset diffuse gastric cancer; since then, the majority of E-cadherin germline mutations have been reported in diffuse type gastric cancer (4–7). Recently, we reported a MET germline mutation as well as E-cadherin germline mutations in the diffuse type of familial gastric cancer (FGC) (6,8). E-cadherin is found predominantly in epithelial cells and plays a pivotal role in maintaining tissue integrity (9). A large number of reports have identified down-regulation of E-cadherin expression in human carcinomas, and E-cadherin function is lost during the development of most epithelial cancers. Indeed, it is thought that loss of E-cadherin function in cancer cells probably plays an important role in tumor development and progression (10). However, it is not yet understood how these losses of expression are governed. Just as nucleotide variations in the coding region of a gene can alter protein function, polymorphisms within the 5′-promoter region have been known to change the transcriptional efficiency of a variety of genes (11,12). Recently, two frequent polymorphisms in human cancers have been identified in the promoter region of the E-cadherin gene. The first is a C→A single nucleotide polymorphism (SNP) −160 nt from the transcriptional start site of the E-cadherin gene promoter; transcription of the A allele is 68% less efficient than that of the C allele (13). The second reported promoter variant is a G→A SNP −347 nt from the transcriptional start site of the E-cadherin gene. The original report suggested that this polymorphism had no effect on transcriptional activity (14). In this study, we sought to better understand the mechanisms of altered E-cadherin expression by investigating −347 G→GA polymorphism effects on E-cadherin transcriptional activity.

**Materials and methods**

**DNA isolation from blood samples**

Blood samples of 28 cases from 27 FGC families without germline mutations in the E-cadherin coding sequence (8) and 142 normal control individuals were collected from the Seoul National University Hospital, South Korea. Informed consent was obtained from all participants prior to testing. Twenty-seven Korean families affected with familial gastric cancer were investigated for genotyping of −347G→GA promoter polymorphism of E-cadherin gene. Criteria for family inclusion were at least two first or second degree relatives affected with gastric cancer, at least one of whom was diagnosed with cancer before the age of 50 (8). Out of 27 probands (range 22–69 ages), 12 represented families suffering from diffuse types of gastric cancer, four represented families suffering from intestinal types and histological data for the type of the remaining 11 were not available. The classification of hereditary diffuse gastric cancer or hereditary intestinal gastric cancer was not possible in these families owing to the lack of histological information. The normal control population was randomly selected from 142 healthy Korean individuals. Peripheral blood lymphocytes were isolated from samples using Ficoll-Paque (Pharmacia Biotech, Uppsala, Sweden) according to the manufacturer’s instructions. Total genomic DNA was extracted with the Trizol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer’s instructions.

**DNA analysis of the E-cadherin promoter regions**

We screened each of the above samples for the −347G→GA E-cadherin polymorphism using denaturing high-performance liquid chromatography (DHPLC) (WAVE®, Transgenicom, Omaha, NE). DNA fragments containing...
the promoter region of interest were amplified with the following primers: forward, 5'-CGCCCGCTGTGCTCCTAC-3', reverse, 5'-GGCCAAGCCTCATTAGGAC-3'. PCR amplification for DHPLC analysis was carried out in a volume of 25 µl containing 100 ng genomic DNA, 10 pmol of each primer, 0.25 mM each dNTP, 0.5 U of Taq polymerase and the provided reaction buffer (GeneCraft, Münster, Germany). Reactions were carried out in a programmable thermal cycler (MWG Biotech AG, Ebersberg, Germany) as follows: denaturation for 5 min at 94°C, followed by five cycles of 94°C for 30 s, 65°C for 30 s, 72°C for 1 min, then 30 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 1 min, followed by five cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, then extension for 10 min at 72°C. DHPLC was performed as described previously (15,16). For heteroduplex formation, PCR products were denatured at 95°C for 5 min followed by gradual cooling to 25°C over a period of 1 h. All samples were investigated by DHPLC and direct sequencing. Direct sequencing was carried out using a Big-dye terminator cycle sequencing kit and an ABI 3100 DNA sequencer (Perkin-Elmer, Foster, CA).

PCR-RFLP

To investigate the frequency of the −160C→A polymorphism, we performed PCR-RFLP analysis. PCR primers and conditions were the same as above. PCR products were digested with Hinfl, and separated on a 3% agarose gel. The A allele yielded two fragments (369 and 79 bp), whereas the C allele was visualized as a single band (448 bp). The results were confirmed by direct sequencing as above.

Transient transfection and dual luciferase reporter assay

To examine the potential effect of the −347G→A polymorphism on E-cadherin gene transcription, a 794 bp promoter region of the E-cadherin gene (from −647 to −1147) carrying either the G or GA allele was inserted upstream of the firefly luciferase gene in the pGL3 Enhancer plasmid vector (Promega, Madison, WI). The G and GA alleles were amplified from DNA samples taken from FGC patients, digested with KpnI and BglII, and cloned into the vectorless pGL3 enhancer plasmid vector. Three different luciferase reporter plasmids were generated: pGL3-G (containing the G allele), pGL3-GA (containing the GA allele) and pGL3-control (Promega), which contains SV40 promoter and enhancer sequences. Each construct was confirmed by sequencing. We performed transient transfections in CV-1, HeLa, SNU-719, AGS and KatoIII cells obtained from Korean Cell Line Bank. HeLa, SNU-719, AGS and KatoIII cells were cultured in RPMI 1640 supplemented with 10% FBS, 100 U/ml penicillin, 100 mg/ml streptomycin and 0.25 mM each dNTP, 0.5 U of Taq polymerase and the provided reaction buffer (GeneCraft, Münster, Germany). Reactions were carried out in a programmable thermal cycler (MWG Biotech AG, Ebersberg, Germany) as follows: denaturation for 5 min at 94°C, followed by five cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, then 30 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, followed by five cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, then extension for 10 min at 72°C. DHPLC was performed as described previously (15,16). For heteroduplex formation, PCR products were denatured at 95°C for 5 min followed by gradual cooling to 25°C over a period of 1 h. All samples were investigated by DHPLC and direct sequencing. Direct sequencing was carried out using a Big-dye terminator cycle sequencing kit and an ABI 3100 DNA sequencer (Perkin-Elmer, Foster, CA).

EMSAs

EMSAs were performed using the Gel Shift Assay System (Promega). Complementary oligonucleotides corresponding to the E-cadherin promoter sequence were synthesized (Bioneer, Seoul, South Korea) as follows (bold letters indicate polymorphism): −347G (containing the G allele), 5'-GGGTGAACAGATGGAGCCCTCCTCAACAA-3', −347GA (containing the GA allele), 5'-GGGTGAACAGATGGAGCCCTCCTCAACAA-3'. The oligonucleotide pairs were annealed and labeled with [γ-33P]ATP (Amersham Biosciences, Buckinghamshire, UK). Binding reactions were carried out with HeLa nuclear extracts in accordance with the manufacturer’s instructions. One microlitre of 32P-labeled probe was incubated with 5 µg of HeLa nuclear extract for 20 min at room temperature. DNA–protein binding specificity was tested by competition assays in which the binding reactions were pre-incubated with 10-, 50- and 100-fold excesses of unlabeled specific or non-specific competitor oligonucleotides prior to the addition of the labeled probe. After the binding reaction was complete, the DNA–protein complexes were resolved by electrophoresis in a 4% non-denaturing acrylamide gel. After electrophoresis, the gels were transferred onto 3 M Whatman paper, dried and autoradiographed.

Statistical analysis

The χ2 test for association was used to assess the differences in genotype distribution. The genotypic-specific risks were estimated as odds ratios (OR) with associated 95% confidence intervals (CI) by unconditional logistic regression (17). The observed genotypes frequencies were compared using a χ2 test to determine if they were in Hardy–Weinberg equilibrium. All tests were performed with the STATISTICA software package (StatSoft Inc., Galvaniweg, UK). A P value < 0.05 was considered statistically significant.

Results

Effect of the −347G→A polymorphism on promoter activity

To examine the effect of the −347G→A promoter region polymorphism on transcription of the E-cadherin gene, we measured promoter activity with a Dual Luciferase Reporter Assay System (Promega) and compared the activities of the −347GA and −347GG alleles by transient transfection assay in CV-1, HeLa, SNU-719, AGS and KatoIII cells. As shown in Figure 1, significantly lower luciferase activities were generated by the pGL3-GA construct as compared with the pGL3-G construct. In CV-1 cells, the GA allele decreased the transcriptional efficiency by 10-fold (P = 0.000261) compared with the G allele. Similar results were obtained in HeLa, SNU-719, AGS and KatoIII cells (12-, 8-, 9-, 13-fold decrease, respectively).

Effect of the −347G→A polymorphism on the binding activity of nuclear factors

to determine whether the −347G→A polymorphism affects the binding activity of nuclear factors, synthetic −347G and −347GA oligonucleotides were incubated with HeLa cell nuclear extracts and subjected to EMSA. The −347GA oligonucleotide showed weak DNA–protein binding, whereas the −347GGA oligonucleotide showed stronger DNA–protein binding (Figure 2). To verify the DNA–protein complex, competition assays were performed with specific and non-specific oligonucleotides (Figure 2). When a −347GGA oligonucleotide was used to compete with a −347GGA oligonucleotide, it totally disrupted the −347GGA oligonucleotide binding with nuclear protein. However, when a −347GA oligonucleotide was used to compete with a −347GA oligonucleotide, it was not as effective as a −347GGA oligonucleotide in disrupting oligonucleotide binding with nuclear protein.

Allele frequencies in FGC samples and normal controls

to determine whether there is a correlation between the promoter polymorphisms and FGC, we screened a 448 bp region (−529 to −82 from the transcriptional start site) of the E-cadherin promoter in 28 cases of FGC and 142 normal controls using PCR-RFLP or DHPLC (Figure 3). We identified the previously reported −160C→A and −347G→A polymorphisms (13,14), and noted a positive association between the −347GA allele and FGC. Eleven (39.4%) of 28 FGC samples were heterozygous at this locus, as compared with 39 (27.5%) of 142 normal controls. Individuals with the E-cadherin genotype (−347GA/−347GA) and the E-cadherin (−347GA/GA) and FGC (Table I).

The distribution of genotypes was in Hardy–Weinberg equilibrium. In contrast, we did not identify a positive or negative association between the −160C→A polymorphism and FGC (Table I).

Discussion

In the present study, we focused on the effect of the −347G→A E-cadherin polymorphism on transcriptional activity. Several major cis-acting elements have been identified within a short section of the proximal promoter of the E-cadherin gene, including two E boxes, a CAAT box, and a
The GC-rich box SP1 binding site (14,18). The E-cadherin gene promoter thus exhibits a modular structure, suggesting that the strict control of epithelium-specific E-cadherin expression might result from interactions among the various regulatory elements (19). Our results demonstrate that the −347 SNP has a significant effect on transcriptional activity in transient transfection experiments. We performed transient transfections in CV-1, HeLa, SNU-719, AGS and KatoIII cells, because it was...
reported previously that the $-347 \text{G} \rightarrow \text{GA}$ promoter polymorphism of $E$-cadherin gene had no effect on transcriptional activity in CV-1 cells (14). In contrast to previous reported results, our study showed that in CV-1 cells, the GA allele of this polymorphism decreased the transcriptional efficiency by 10 fold ($P \leq 0.000261$) compared with the G allele (Figure 1). Similar results were obtained in HeLa, SNU-719, AGS and KatoIII cells. The molecular mechanism of this difference may relate to differences in the affinity of DNA-binding proteins to the two alleles of the $E$-cadherin promoter. We searched for putative transcriptional factors that might bind with the $-347$ SNP, using the Ds gene software package (Accelrys, San Diego, CA). We identified four putative transcription factors (Site_C2, ZESTE_CS, T-Ag-SV40.3, T-Ag-EP) with similarities to sequences near the $E$-cadherin $-347\text{G} \rightarrow \text{GA}$ promoter polymorphism. These putative transcription factors are not well characterized, and their in-depth study may be a target of future work. To investigate binding between the alleles and nuclear factors in general, we performed EMSA, which revealed that the $-347\text{GA}$ allele bound nuclear factors more weakly than the $-347\text{G}$ allele. In competition assay, the $-347\text{G}$ allele was able to disrupt $-347\text{GA}$-protein binding, whereas the $-347\text{GA}$ allele was less able to disrupt nuclear protein binding to the $-347\text{G}$ allele. Although further work will be necessary to investigate the exact molecular mechanism by which activity of $E$-cadherin is affected by the allelic variation, our results suggest that the $-347\text{GA}$ polymorphism may negatively impact transcription factor binding, leading to a decrease in $E$-cadherin expression. Lastly, we examined whether the $-347\text{G} \rightarrow \text{GA}$ promoter polymorphism of the $E$-cadherin gene was associated with FGC. Individuals with the $E$-cadherin genotype ($-347\text{G}/\text{GA}$ heterozygous or GA homozygous) had an increased risk for FGC. In the case of the $-160\text{C} \rightarrow \text{A}$ polymorphism, several reports have investigated the correlation between the $-160\text{C} \rightarrow \text{A}$ polymorphism of the $E$-cadherin and gastric cancer (2,17,20). However, the correlation between the $-160\text{C} \rightarrow \text{A}$ polymorphism of the $E$-cadherin and gastric cancer is still controversial. We found that the genotype frequency of $-160\text{C} \rightarrow \text{A}$ polymorphism did not differ between the normal controls and FGC patients (Table I). However, the sample number was too small to determine the statistical significance of these differences. In the future, larger population studies will be required to confirm whether these variants increase the risk of cancer in Korean and other ethnic groups.

In summary, we investigated the importance of the $-347\text{G} \rightarrow \text{GA}$ polymorphism in the promoter region of the $E$-cadherin gene. The GA allele was associated with significant suppression of $E$-cadherin transcription in CV-1, HeLa, SNU-719, AGS and KatoIII cells. Additionally, EMSA revealed that the GA allele had a weak transcriptional factor binding strength compared with the G allele. Therefore, it seems that

---

Table I. Genotype and allele frequencies of $E$-cadherin polymorphisms

<table>
<thead>
<tr>
<th>Polymorphisms</th>
<th>Normal ($n = 142$)</th>
<th>FGC ($n = 28$)</th>
<th>OR$^a$ (95% CI$^b$)</th>
<th>$\chi^2$</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-347$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>103 (72.5%)$^c$</td>
<td>16 (57.1%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G/GA</td>
<td>39 (27.5%)</td>
<td>11 (39.3%)</td>
<td>1.815 (0.7-4.2)</td>
<td>6.97</td>
<td>0.03059</td>
</tr>
<tr>
<td>GA</td>
<td>0 (0%)</td>
<td>1 (3.6%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>110 (77.5%)</td>
<td>21 (75%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-160$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/A</td>
<td>31 (21.8%)</td>
<td>6 (21.4%)</td>
<td>1.01 (0.3-2.7)</td>
<td>1.65</td>
<td>0.43726</td>
</tr>
<tr>
<td>A</td>
<td>1 (0.7%)</td>
<td>1 (3.6%)</td>
<td>5.2 (0.3-87.1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Odds ratio compared with homozygous individuals (GG or CC).

$^b$95% CI; OR and 95% CIs were calculated by logistic regression with the GG or CC genotype as the reference group.

Allele frequency.

---

Fig. 3. $E$-cadherin $-347 \text{G} \rightarrow \text{GA}$ polymorphism. (A) $-347\text{G}$ (containing the G allele); (B) $-347\text{GA}$ (containing the GA allele). The underline denotes the SNP site; arrow indicates DHPLC chromatogram and matched sequencing chromatogram. See online supplementary material for colour version of this figure.
A functional polymorphism (−347G→GA) of E-cadherin gene

−347G→GA polymorphism may affect the expression of E-cadherin, possibly increasing the cancer risk.

Supplementary material
Supplementary material can be found at: http://www.carcin.oupjournals.org/

Acknowledgements
This work was supported by a research grant from the National Cancer Center, Korea, grant HMP-99-M-03-0001 from the Ministry of Health and Welfare, and the BK21 project for Medicine, Dentistry and Pharmacy.

References

Received August 20, 2003; revised November 16, 2003; accepted December 19, 2003