A Functional Common Polymorphism in a Sp1 Recognition Site of the Epidermal Growth Factor Receptor Gene Promoter

Wanqing Liu,1 Federico Innocenti,1,3,5 Michael H. Wu,1 Apurva A. Desai,1,3 M. Eileen Dolan,1,3,4,5 Edwin H. Cook, Jr.,3,5 and Mark J. Ratain1,3

Departments of Medicine and Psychiatry, Pediatrics, and Human Genetics; Committees on Clinical Pharmacology and Pharmacogenomics and Cancer Biology; and Cancer Research Center, University of Chicago, Chicago, Illinois

Abstract
The epidermal growth factor receptor (EGFR) plays a prominent role in cell growth and development. Its regulation in humans is complex and incompletely understood. In this study, 12 new polymorphisms were discovered in the 5′-regulatory region of EGFR gene and 2 common single nucleotide polymorphisms (−216G/T and −191C/A) were found in the essential promoter area, one of which is located in a Sp1 recognition site (−216). Transient transfection in human cancer and primary cell lines showed significantly different promoter activity between the two most common haplotypes (−216G/−191C and −216T/−191C). The replacement of G by T at position −216 increases the promoter activity by 30%. A transient transfection assay in the Sp1-deficient cell line (Schneider cell line 2) showed a strong dependence of EGFR promoter activity on Sp1 and confirmed the effect of the aforementioned polymorphisms. Electrophoretic mobility shift assay also showed a significantly higher binding efficiency of nuclear protein or pure Sp1 protein to the T allele compared with the G allele. We then investigated the allelic imbalance of EGFR transcription in fibroblast cell lines with heterozygous genotype at −216G/T but C/C homozygous genotype at −191C/A. The expression of mRNA carrying T-C by A-T at position −216 shows a significantly greater imbalance of EGFR transcription in fibroblast cell lines with heterozygous genotype at −216G/T but C/C homozygous genotype at −191C/A. The expression of mRNA carrying T-C haplotype was significantly stronger compared with that of G-C haplotype (P < 0.02). Thus, we successfully showed that a common polymorphism in the EGFR promoter was associated with altered promoter activity and gene expression both in vitro and in vivo. Our findings have implications for cancer etiology and therapy and may also be relevant to the inherited susceptibility of other common diseases. (Cancer Res 2005; 65(1): 46-53)

Introduction
The human epidermal growth factor receptor (EGFR) gene is located in chromosome 7p12.1-12.3. Its product, a 170-kDa transmembrane glycoprotein, plays a critical role in the signal transduction pathway for cell proliferation, differentiation, and survival. Overexpression of EGFR is found in ~30% of human primary tumors and has been significantly associated with disease stage, prognosis, survival, and response to chemotherapy (1). EGFR is also a possible genetic risk factor for cancer susceptibility. In transgenic mice, overexpression of the EGFR initiates formation of oligodendrogloma and breast cancer (2, 3). In humans, a (CA)n polymorphism in intron 1 of EGFR was recently associated with breast cancer risk in young women (4). Because of its important function in cancer, EGFR is now an attractive target for treatment and prevention of cancer. Several anticancer agents inhibiting the EGFR phosphorylation or blocking the ligand binding are being tested currently in clinical trials, and two anti-EGFR agents, cetuximab and gefitinib, were recently approved by the U.S. Food and Drug Administration for the treatment of colorectal and lung cancer, respectively.

However, the regulation of EGFR is not yet completely understood. To date, several features of the 5′-regulatory region of EGFR have been described, including a TATA-less, CAAT-less, high GC content promoter with multiple transcriptional start sites (5, 6); an intronic enhancer element in hormone-independent cells and other two cooperative enhancer regions around exon 1 (7, 8); and a variety of trans and cis elements, such as epidermal growth factor–responsive DNA binding protein-1, p53, p63, Sp1, vitamin D–responsive element, and estrogen-responsive element (5, 6, 9–20). Furthermore, multiple Sp1 binding sites have been discovered, and Sp1 is deemed necessary for EGFR promoter activity (5, 9, 21, 22). However, only limited data exist regarding DNA variants in the 5′-regulatory region of EGFR. Recently, a microsatellite sequence ~1.5 kb downstream of exon 1 has been proposed to regulate EGFR expression (23, 24). Thus far, no regulatory variations in the promoter or enhancer regions of EGFR with functional consequence have been reported. To further understand the regulation of EGFR, we resequenced an ~4-kb area in the 5′-regulatory region of EGFR in Caucasian, African American, and Asian populations. Furthermore, we characterized the function of two common polymorphisms in the EGFR promoter region.

Materials and Methods
Discovery of Single Nucleotide Polymorphisms in EGFR Regulatory Region. DNA samples from either the Lymphoblastoid Cell Line Core in University of Chicago or the Coriell Cell Repository were used for resequencing. The samples included 24 African Americans, 22 Caucasians, and 23 Asians. For single nucleotide polymorphism (SNP) discovery, PCR was used to amplify the ~4.0-kb fragment containing the upstream and downstream enhancer, promoter, exon 1, and part of intron 1. Primers were designed according to the reference sequence AF288738 from Genbank using Primer Premier version 5.00 (Premier Biosoft International, Palo Alto, CA). Primer sequences were EGFRF: 5′-GATGACCTACAC-CCAGGAT-3′ and EGFRR: 5′-GAGGCTAAGTGC-CACTGC-3′, EGFR2F: 5′-TCGAGCTTAGAGCACACCG-3′ and EGFR2R: 5′-GAGACCTGGATATGC-3′, EGFR3F: 5′-AAATTTACTCCAGCCGACC-3′ and EGFR3R: 5′-CGCCCTTACCTTTCTTCCC-3′, EGFR4F: 5′-CCCTGACTCGGTCA-GTATT-3′ and EGFR4R: 5′-AAGAAAGTTGGACCGGTC-3′, EGFR5F: 5′-CGCTTCTCCTGGTCGTTGG-3′ and EGFR5R: 5′-AGACAAAGTCTCCAGGCTCC-3′, EGFR6F: 5′-GCCAGGCTTCAAGTAAG-3′ and EGFR6R:
5'-GGAGAACTTTGCTGAGCC-3', EGRF7F: 5'-CCCTCGTCTGCTCA-3' and EGRF7R: 5'-AGTGATCCTCAATGCTG-3', EGRF8F: 5'-GGATGAGAACAGTGGTTCCC-3'; EGRF8R: 5'-GACACAAATGGGAGGAGAA-3', and EGRF8R: 5'-TGAAGAAGCTTTGGAGAAAGG-3' and EGRF9R: 5'-CATGTCGGCAACACAAACA-3'. After PCR, products were purified and subsequently directly sequenced from both ends.

**SNP Location, Allele Frequency, Haplotypes, and Linkage Disequilibrium.** Position of SNPs is reported relative to the initiator ATG (with A as +1). The common and rare alleles were separated by a slash (the common allele is always put to the left of the slash; Table 1). Allele frequency of each SNP was calculated by allele counting among each ethnic group. Pairwise linkage disequilibrium (LD) and haplotypes were calculated or estimated by using Arlequin 2.000 (http://anthro.unige.ch/arlequin). D' and r² were generated and plotted through the software, LD plotter (http://innateimmunity.net/IIPGA2/Bioinformatics/).

**Genotyping of −216G/T, −191 C/A, and Intron 1 (CA)n Polymorphisms.** Restriction enzymes were used to genotype the two SNPs in the EGRF promoter. The G-to-T change at position −216 is recognized by BseRI, whereas the C-to-A at −191 abolishes the cutting site of SacII. Genotyping of these two polymorphisms was done by digesting 10-μl PCR product using 8 units BseRI or 10 units SacII (New England Biolabs, Beverly, MA), respectively, at 37°C for 2.5 hours. Digestion reaction was then inactivated at 65°C for 15 minutes, separated in 2% agarose gel, and visualized by ethidium bromide staining. The intron 1 (CA)n polymorphism was genotyped as described previously (25).

**Construction of EGRF Promoter Reporter Vectors.** Previous deletion mapping studies have shown that −500 bp 5'-flanking sequence (−848 to −16) upstream of the initiator ATG has the essential promoter function, whereas the fragment −384 to −16 has full promoter activity (Fig. 1; refs. 6, 9, 21). Therefore, this fragment containing the −216G/T and −191C/A polymorphisms was chosen to construct the luciferase reporter gene vectors. PCR was also done to generate a 525-bp amplicon (−418 to +107) using 5'-CCACCGGTACCCGCGCGCCTGGGCTTG-3' as the forward primer and 5'-CGGCGAGACACGCCCTTACCTTT-3' as the reverse primer. This fragment contained a SacI site at position −16 (Fig. 1). To facilitate the subcloning, the forward primer was designed to contain a KpnI site (modified nucleotides were indicated by italics in the forward primer sequence listed above). This fragment was amplified from the individuals with specific haplotypes of −216G/T and −191C/A by PCR using the Proofstart DNA polymerase (Qiagen, Hilden, Germany), which is modified for high-fidelity DNA amplification. The fragment was then digested by KpnI and SacI and a 397-bp product (−412 to −16) was cloned into the KpnI/SacI site of pGL3-basic vector containing the firefly luciferase (Promega, Madison, WI) to construct vectors designated as pGL3EGFRIIuc (Fig. 1). Three haplotypes [−216G-191C (G-C, *1), −216G-191A (G-A, *2), and −216T-191C (T-C, *3)] were successfully amplified and cloned. The rare haplotype −216T-191A (T-A, *4) was constructed by ligating the T-fragment and the A-fragment from the DraIII-digested T-C and G-A haplotypes (Fig. 1). Plasmid DNA was prepared and all clones were sequenced to exclude any PCR errors and to ensure the orientation of the haplotypes prior to transfection.

**Cell Culture.** Human breast cancer cell lines MDA-MB-231 and MCF-7 were grown in RPMI 1640 and the human embryonic kidney 293 (HEK-293) cells were grown in DMEM. Both media were supplemented with 2 mmol/L glutamine and 10% fetal bovine serum in a humidified atmosphere (5% CO₂, 95% air). Drosophila melanogaster Schneider cell line 2 (SL-2) was obtained from the American Type Culture Collection (Manassas, VA) and maintained Schneider's Drosophila medium supplemented with 10% fetal bovine serum at 23°C and atmospheric CO₂. Human fibroblast cell lines were obtained from the Coriell Cell Repository and grown in Eagle's MEM with Earle's salts supplemented with 2 mmol/L glutamine and 15% fetal bovine serum at 37°C and atmospheric CO₂. All media and supplements for cell culture were purchased from Invitrogen (Carlsbad, CA).

**Transient Transfection.** The pGL3EGFRIIuc reporter vectors carrying each of the four haplotypes (*1 to *4) were transfected into MDA-MB-231, MCF-7, HEK-293, or SL-2 cells, respectively, to compare the relative expression of the luciferase gene. Transient transfection of MDA-MB-231 and HEK-293 cells was done with LipofectAMINE 2000 (Invitrogen), whereas transfection of MCF-7 and SL-2 was done with FuGene 6 transfection reagent (Roche Molecular Biochemicals, Mannheim, Germany) according to the manufacturer's instructions. Human cells were cotransfected by each pGL3EGFRIIuc reporter vector together with pBl-TK reporter vector (Promega). The SL-2 cells were transfected with pGL3EGFRIIuc alone or together with pPac-Sp1 [(Sp1 expression vector pPac-Sp1 was kindly provided by Drs. Robert Tjian (University of California, Berkeley, CA) and Erin G. Schuetz (St. Jude Children's Research Hospital, Memphis, TN)]. For detection, cells were washed twice with PBS and lysed and luciferase activity was determined using the dual luciferase reporter assay system (Promega) according to the manufacturer's instructions. Experiments were done in triplicate and were repeated at least twice. To control the transfection efficiency, activity of firefly luciferase was normalized to either

### Table 1. Number and frequency of rare alleles of SNPs discovered from EGRF regulatory region

<table>
<thead>
<tr>
<th>SNP</th>
<th>Surrounding sequence</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>African Americans</td>
</tr>
<tr>
<td>−1433C/T</td>
<td>ctcagagggtCgagaccacgg</td>
<td>4.2</td>
</tr>
<tr>
<td>−1298G/A</td>
<td>agagacagTggtgtgaacc</td>
<td>4.2</td>
</tr>
<tr>
<td>−1247G/A</td>
<td>gagaggagggTggtgtgtga</td>
<td>10.4</td>
</tr>
<tr>
<td>−1255G/A</td>
<td>catgcactagaggacattt</td>
<td>0</td>
</tr>
<tr>
<td>−759C/A</td>
<td>ctcagccggAaggegctccg</td>
<td>14.6</td>
</tr>
<tr>
<td>−646G/A</td>
<td>tcaaatgctgTcaagatttg</td>
<td>4.2</td>
</tr>
<tr>
<td>−540G/A</td>
<td>cgccggacccGtcacagg</td>
<td>0</td>
</tr>
<tr>
<td>−482C/A</td>
<td>tcaggcaccGctccctcc</td>
<td>6.3</td>
</tr>
<tr>
<td>−216G/T</td>
<td>acaegctcCcGccccccgac</td>
<td>29.2</td>
</tr>
<tr>
<td>−191C/A</td>
<td>tgacgcgcggCgcgcgag</td>
<td>0</td>
</tr>
<tr>
<td>169G/T</td>
<td>ggcagcgcggGgacgctc</td>
<td>12.5</td>
</tr>
<tr>
<td>2028G/A</td>
<td>ctcttagcCggtgtcagcc</td>
<td>4.2</td>
</tr>
</tbody>
</table>

**NOTE:** SNP is in uppercase. SNPs are named relative to the initiator codon with A of the ATG codon as +1 and followed by common/rare allele.

www.aacrjournals.org 47 Cancer Res 2005; 65: (1). January 1, 2005
Renilla luciferase of pRL-TK vector (human cell lines) or the total cellular protein level measure by the Bradford method (SL-2, Bio-Rad, Hercules, CA).

Electrophoretic Mobility Shift Assay. Nuclear proteins were extracted from MDA-MB-231 cells using the NE-PER nuclear and cytoplasmic extraction reagents according to the manufacturer's protocol (Pierce, Rockford, IL). Pure human Sp1 protein was purchased from Promega. The probes and competitors for −216G, −216T, and the Sp1 consensus binding sequence (Promega) were 5′-GCAGCTCCGCCGCCAGCAGTTGT-3′, 5′-GCAGCC-TCCGCCGCCAGCAGTTGT-3′, and 5′-ATTGACCCCGCCGCAGCCGC-3′, respectively. Probes were synthesized as single strands and end labeled by biotin. Identical unlabeled oligonucleotides with same sequences were used as competitors. dsDNA was made and electrophoretic mobility shift assay was done using LightShift Chemiluminescent EMSA kit (Pierce) according to the manufacturer's instruction.

Allele-Specific Transcription and Real-time PCR. Total RNA was extracted from MDA-MB-231, MCF-7, HEK-293, and 10 human fibroblast cell lines. Fibroblast cells were selected based on genotypes (G/T heterozygotes, A/G homozygotes at −191). cDNA of each sample was then obtained through reverse transcription of 1 μg total RNA using SuperScript II reverse transcriptase with random hexamer primers according to the manufacturer's protocol (Invitrogen).

cDNA from MDA-MB-231, MCF-7, and HEK-293 was then used to perform real-time PCR to evaluate the expression of EGFR gene with SYBR Green I method (Bio-Rad). β-actin gene was included as internal control. Primer sequences were for EGFR: 5′-GTCGGCATCCGCTTGGTCTGC-3′ (forward) and EGFR-rF: 5′-CAAGAAAGGGTGTAACGCAACTA-3′ (reverse) for cDNA. The reverse primer for cDNA amplification was particularly designed into the 5′-CAGGGTTCGCTGCCTCCGGCTC-3′ region of EGFR. Nested PCR products were then purified by treatment with ExoI and shrimp alkaline phosphatase (Roche Molecular Biochemicals). Purified PCR product was then used to perform single base extension to genotype the −216G/T polymorphism with primer EGFR-SBE (5′-CTCACCAGGTCGGGAGG-3′) and the SNaPshot kit (Applied Biosystems, Foster City, CA). The ratio of T allele and G allele TcDNA/GcDNA (represented by the value of peak height) in the cDNA was normalized to the TcDNA/GcDNA ratio in the corresponding DNA. The 95% confidence intervals for the relative ratios [R = (TcDNA / GcDNA) / (TcDNA / GcDNA)] from 10 samples were calculated using GraphPad Prism version 3.00 for Windows (GraphPad Software, San Diego, CA; http://www.graphpad.com).

Results

Single Nucleotide Polymorphism Discovery. By sequencing the 4-kb EGFR 5′-regulatory region, including the promoter and enhancers, 12 new SNPs were identified (Table 1). Five SNPs showed relatively high frequency (rare allele frequency ≥10%) in at least one of the three populations screened (Table 1), among which the −216G/T was the most common polymorphism in African Americans (29.2%) and Caucasians (31.8%) populations but not in Asians (7.1%). The African American population had the highest DNA diversity with one SNP in 449 nucleotides on average compared with one only in 674 among Caucasians and one in 1,348 among Asians.

All but 1 (169 G/T) of the 12 SNPs was located in the enhancer or promoter region, reflecting a highly polymorphic regulatory region of EGFR. Two common polymorphisms, −216G/T and −191 C/A, were located in a critical region for the promoter activity where all transcriptional start sites and multiple nuclear protein affinity sites are located (6, 9). Interestingly, the −216 G/T was located in a Sp1 binding site (6, 9) and the −191 C/A was just 4 bp upstream of one of six transcription initiation sites (6). Therefore, these two SNPs may have a significant impact on regulation of EGFR.
Linkage Disequilibrium and Haplotype Prediction. Pairwise LD across the 4-kb region was shown in Fig. 2. The \(-216G/T\) showed a very low level of LD with other SNPs in this region by both \(D'\) and \(r^2\) values. To understand a potential relationship between the newly discovered SNPs and the intron 1 (CA)_n polymorphism located between exon 1 and the downstream enhancer region and previously suggested to have a regulatory function on EGFR expression (23, 24), pairwise LD based on \(P\) values of Fisher’s exact test was calculated using the Arlequin software (the LD plotter cannot calculate the LD including microsatellite data). There was a very low level of LD (\(P > 0.05\), Fisher’s exact test) between \(-216G/T\) and the (CA)_n, whereas the other five relatively common SNPs were in tight LD with the (CA)_n (\(P < 0.01\), Fisher’s exact test; Fig. 2C). This implies that the \(-216G/T\) may have an independent effect on regulating EGFR gene compared with the intron 1 (CA)_n polymorphism.

Haplotypes across all the polymorphisms were predicted and six major ones were listed in Table 2. There is an obvious ethnic difference in haplotype distribution among three populations but with a relatively low haplotype diversity, as a few common haplotypes accounted for >60% of all haplotypes. Effective number of haplotypes was significantly lower in Asians than in other two populations, indicating a lowest haplotype diversity in this population (Table 2). With regard to the two potentially functional
SNPs in the promoter region (–216G/T and –191C/A), three haplotypes (G-C, G-A, and T-C) were observed in our samples. The G-C and T-C were the two common haplotypes in the essential EGFR promoter among all three populations, whereas the G-A haplotype was only found in Caucasians (Table 2).

**Transient Transfection Assay.** Promoter activity of the four haplotypes (*1 to *4) showed similar patterns using either human primary embryonic epithelial cells (HEK-293) or cancer cell lines (MDA-MB-231 and MCF-7) with significantly higher luciferase activity for the T-C haplotype in comparison with the G-C haplotype vector (P < 0.04 for all comparisons; Fig. 3A). This effect was independent of the *EGFR* expression level of the cells (Fig. 3B). The –216 G/T polymorphism had a greater effect on activity than the –191 C/A polymorphism. On average, the G-to-T substitution showed an ~30% increase of promoter activity.

To further confirm the potential cooperative effect of the DNA alteration and Sp1 on promoter activity, transient transfection was also done in the *D. melanogaster* SL-2 in which Sp1 is deficient (26). As a result, cotransfection of pGL3EGFRluc with Sp1 expression vector resulted in ~100-fold induction of promoter activity compared with transfection of pGL3EGFRluc alone (data not shown). Cotransfection of pPac-Sp1 and each of four pGL3EGFRluc constructs showed a significantly lower promoter activity driven by G-C haplotype compared with the T-C haplotype (P < 0.03; Fig. 3A).

**Preferential Protein-DNA Interaction between Nuclear Extracts or Sp1 and –216G/T by Electrophoretic Mobility Shift Assay.** Electrophoretic mobility shift assay was done to test the binding efficiency of nuclear proteins from the MDA-MB-231 cells to each allele-specific probe. Both alleles of –216G/T showed affinity to nuclear proteins at the same position of Sp1 consensus sequence. The band pattern of protein-DNA complex could be changed by excess of competitors. The affinity of nuclear proteins to the T allele was significantly higher than that to the G allele (Fig. 4A). Assays with pure human Sp1 protein produced similar results (Fig. 4B). These data further support our view that the altered *EGFR* promoter activity associated with the –216G/T was a result of differential affinity of transcription factor to the variant site.

**Haplotypes of –216G/T−191C/A Were Associated with EGRF mRNA Expression In vivo.** We selected human fibroblast cells (which express *EGFR*) to evaluate the association between –216G/T−191C/A haplotypes and *EGFR* transcription. According to previous reports, there were multiple transcription initiation sites in the *EGFR* promoter (6, 9), whereas the major site for *in vivo* transcription was at position –260 (Fig. 1; ref. 9). Thus, positions –216 and –191 would be present in most *EGFR* mRNA sequences. We therefore chose 10 cell lines with diplotype G-C/T-C for the two polymorphisms, so that we could potentially detect a difference of expression level between mRNA carrying T-C haplotype and G-C haplotype within the same cell.

As a result, a significant deviation of the average relative ratio from the hypothetical ratio 1:1 was observed (mean R, 1.39 ± 0.12; 95% confidence interval, 1.11-1.67; P < 0.02), demonstrating that *EGFR* mRNA derived from the T-C haplotype was ~40% higher than that from the G-C haplotype. This finding indicates that the –216G/T variant also has a strong impact on *EGFR* transcription in vivo.

### Table 2. Distribution of six major haplotypes of 5′-regulatory region and four haplotypes between –216G/T and –191C/A among populations

<table>
<thead>
<tr>
<th>ID</th>
<th>Major haplotypes of 5′-regulatory region</th>
<th>African American (%)</th>
<th>Caucasian (%)</th>
<th>Asian (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CGGGCGGCGGCG20G</td>
<td>27.1</td>
<td>15.0</td>
<td>34.8</td>
</tr>
<tr>
<td>2</td>
<td>CGGGCGGCGCTCG16G</td>
<td>16.7</td>
<td>17.3</td>
<td>5.3</td>
</tr>
<tr>
<td>3</td>
<td>CGGGCGGCGGCGG16G</td>
<td>12.5</td>
<td>16.8</td>
<td>29.5</td>
</tr>
<tr>
<td>4</td>
<td>CGGGCGGCGGCGG21G</td>
<td>10.4</td>
<td>4.5</td>
<td>8.7</td>
</tr>
<tr>
<td>5</td>
<td>CGGGAGGGCGAT18G</td>
<td>0.0</td>
<td>11.4</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>CGGGCGGCGCTCG20G</td>
<td>0.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Total no. estimated haplotypes</td>
<td>147</td>
<td>66</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Total frequency</td>
<td>66.7</td>
<td>75.0</td>
<td>78.3</td>
</tr>
<tr>
<td></td>
<td>No. effective haplotypes</td>
<td>7.9</td>
<td>8.1</td>
<td>4.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Haplotypes between –216G/T and –191C/A</th>
<th>African American (%)</th>
<th>Caucasian (%)</th>
<th>Asian (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>*1</td>
<td>G-C</td>
<td>70.8</td>
<td>48.2</td>
</tr>
<tr>
<td>*2</td>
<td>G-A</td>
<td>9.2</td>
<td>18.2</td>
</tr>
<tr>
<td>*3</td>
<td>T-C</td>
<td>29.2</td>
<td>34.1</td>
</tr>
<tr>
<td>*4</td>
<td>T-A</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

NOTE: For the major haplotypes, the order of polymorphisms from left to right are –1433C/T, –1298G/A, –1247G/A, –1225G/A, –759C/A, –646G/A, –540G/A, –484C/A, –216G/T, –191C/A, 169G/T, (CA)₆, and 2028G/A.

*Effective haplotypes, reflecting the frequency of common haplotypes among population, was calculated as the reciprocal of the sum of the haplotype frequency squared. From this, Asian population showed a lowest haplotype diversity among three populations.*
In addition to the allelic imbalance, we also evaluated the relative expression of *EGFR* among the above three human cell lines by real-time PCR. Interestingly, the *EGFR* levels among these cells were in agreement with their diplotypes, with a dramatically high level of *EGFR* in MDA-MB-231 cells but ~6-fold less in HEK-293 and the lowest in MCF-7 (Fig. 3B).

**Discussion**

In this study, we discovered 12 SNPs in the *EGFR* 5′-regulatory region. Some of them occurred with a high frequency, implying a highly polymorphic regulatory region of human *EGFR*. We subsequently showed that a common polymorphism in the *EGFR* essential promoter region was associated with altered promoter activity and gene expression both in vitro and in vivo. To our knowledge, this is the first report describing SNPs in the *EGFR* regulatory region, with evidence for a functional cis-acting common polymorphism in the promoter region. SNPs usually modulate gene transcription by interacting with trans-acting elements (27–31). The *EGFR* regulatory region, like that of many other housekeeping genes, has a high GC content and multiple transcriptional start sites but no TATA box. For such genes, the promoters are often activated by Sp family proteins (32, 33). The pivotal role of Sp1 in *EGFR* promoter activity has been well described (5, 9, 21, 22). The presence of Sp1 in the promoter, whereas –191C/A is 4 bp upstream of one of six transcriptional start sites (6, 9). Therefore, it would be not surprising that these polymorphisms might give a modification onto the promoter activity. On the other hand, *EGFR* intron 1 is >120 kb, suggesting that the sequences around exon 1 may have a separate effect on gene regulation. Through LD analysis, –216G/T is not linked to other polymorphisms in this region, including the (CA)ₙ polymorphism in intron 1, indicating that the function of –216G/T would be independent of any effect of this intronic microsatellite. Through the studies here, our data confirmed the requirement of Sp1 for *EGFR* transcription and suggested an effect of –216G/T on *EGFR* regulation. This effect was also independent of cell type (cancer cell lines or transformed human primary cells) and *EGFR* expression level. However, it is still possible that other functional cis-acting elements upstream or downstream of the 5′ sequences would also cooperate with the –216G/T in vivo. More extensive studies of *EGFR* may provide additional insight into the functionality of the –216 SNP.

Previous studies by Gebhardt et al. (23) have shown that the dinucleotide (CA)ₙ polymorphism in the intron 1 of *EGFR* (near the downstream enhancer), ranging from 14 to 21 repeats, seemed to regulate *EGFR* expression. The longer allele with 21 repeats showed an 80% reduction of gene expression compared with the shorter allele with 16 repeats (23, 24). Significant ethnic differences in allelic frequency of this polymorphism were reported recently (25). However, our data suggest that this polymorphism is in relatively high LD with five other common SNPs in the 5′-regulatory region (Fig. 2C) and thus may not be functional by itself. More importantly, the previous articles demonstrating a functional effect of the (CA)ₙ polymorphism were flawed in that the authors used a methodology (heteroduplex analysis of a 4.5-kb fragment) that was unable to detect the variants reported here. Furthermore, although there is no LD between –216G/T and (CA)ₙ, there is a significant enrichment of longer alleles (corresponding to lower *EGFR* expression) on the chromosomes carrying –216G-191C haplotype in our samples. For example, among African American population, 34.6% longer alleles (n ≥ 20) were clustered with G-C haplotype compared with only 2.9% with T-C haplotype. Among Caucasians and Asians, these numbers were 19.9% versus 10.1% and 43.5% versus 0%, respectively (overall χ² = 12.3; df = 1; P < 0.001). Although

---

**Figure 3.** A, transient transfection of pGL3EGFRluc (*1 to *4) in MDA-MB-231, MCF-7, HEK-293, and SL-2 cells. For human cell lines, pGL3EGFRluc (1.6 μg) was cotransfected with pRL-TK vector (160 ng). For SL-2 cells, pGL3EGFRluc (300 ng) was cotransfected with pPac-Sp1 vector (100 ng), and relative expression of 200 light units of luciferase activity/mg total protein/mL was set to 1. Significant difference of promoter activity was observed between G-C and T-C haplotype of –216G/T-191C/A (all Ps < 0.04). B, relative expression of *EGFR* among MDA-MB-231, MCF-7, and HEK-293 cell lines and corresponding genotypes of –216G/T and –191C/A polymorphisms. *EGFR* mRNA level was normalized to 1,000 copies of β-actin gene. Experiments were repeated thrice. Columns, mean; bars, SE.
our data support the hypothesis that the −216 SNP has functional importance, studies in intact cellular systems will provide more credible evidence to support this suggestion, as exemplified by a recently presented association study of the (CA)n polymorphism (34). Studies are ongoing with fibroblast cell lines from unrelated individuals, which will permit analysis of the relationship of the polymorphisms of interest to various phenotypes, including susceptibility to EGFR inhibitors.

Our findings have several implications. First, EGFR overexpression has been associated with adverse disease stage, prognosis, survival, and response to chemotherapy in a variety of human tumors. Although the overexpression of EGFR has generally been correlated with gene amplification, the level of EGFR is still primarily regulated by the abundance of its mRNA (35–37). Therefore, variations in both cis- and trans-acting elements would be relevant to the variable expression of EGFR. In our study, the −216T was associated with a 40% increase in EGFR expression in vivo compared with −216G. In addition, the −216T would be expected to result in a stronger correlation between Sp1 and EGFR expression, potentially resulting in a different signaling network. Thus, it is likely that −216G/T may at least partly contribute to the variability of EGFR expression in malignant cells and may potentially influence the cell’s dependence on EGFR.

Second, EGFR expression seems to manifest interindividual variability, which was correlated with −216G/T genotype, EGFR is now an attractive target for treatment and prevention of cancer. Studies of EGFR-targeting cancer therapies have shown encouraging results with three EGFR inhibitors: cetuximab, a monoclonal antibody with affinity to the extracellular domain of EGFR, and two small EGFR-specific tyrosine kinase inhibitors, gefitinib and erlotinib (1, 38–42). Two recent studies have shown that somatic mutations in the EGFR tyrosine kinase domain were associated with clinical response of non–small cell lung cancer to gefitinib (43, 44). A more recent study found a similar association with erlotinib as well (45). These gain-of-function mutations were deemed to structurally alter the catalytic pocket of the tyrosine kinase and thus enhance its sensitivity to both ATP and its competitors like gefitinib. Functional assays confirmed that these mutants could activate an antiapoptotic pathway (46).

However, it is likely that other factors also contribute to the variability in response to EGFR inhibitors. The identified somatic mutations were not able to account for all responders with lung cancer. Putting the data from the three studies together, only 81% of lung cancer patients taking gefitinib or erlotinib and experiencing partial responses or marked clinical improvement harbored EGFR somatic mutations (45). Furthermore, not all gefitinib-sensitive patients have mutated EGFR (44). Hence, in spite of strong association, the somatic mutations within EGFR are neither sufficient nor necessary for the drug response. Meanwhile, it was recently suggested that the mutations are not relevant to other tumor types (47), although there is evidence of activity in other diseases such as colon cancer and brain tumors (48). It is also unclear why some patients are more susceptible to somatic mutations in EGFR, with the interethnic differences being of particular interest, possibly suggesting a genetic basis for this susceptibility, either in EGFR or in other genes. Additionally, the occurrence of skin rash has also been associated with antitumor response (49, 50), suggesting that both antitumor response and toxicity to EGFR inhibitors may have a common genetic etiology. This hypothesis is supported by the preliminary results of Perea et al. (34). Additional pharmacogenetic studies are clearly warranted, particularly in conjunction with large phase III studies that include EGFR inhibitors in non–small cell lung cancer.

Acknowledgments

Received 5/26/2004; revised 10/19/2004; accepted 10/27/2004.

Grant support: Pharmacogenetics of Anticancer Agents Research Group (http://pharmacogenetics.org), NIH/National Institute of General Medical Sciences grant U01GM61393; William F. O’Connor Foundation; and PharmGKB (http://pharmgkb.org/), NIH/National Institute of General Medical Sciences grant U01GM61374.

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

We thank Drs. Robert Tjian and Erin G. Schuetz for generously providing the pPac-Sp1 vector, Donna L. Fackenthal for kind help on experiments of single base extension, and Deborah L. Stoit for assistance.

Cancer Res 2005; 65: (1). January 1, 2005 52 www.aacrjournals.org

Downloaded from cancerres.aacrjournals.org on April 20, 2017. © 2005 American Association for Cancer Research.
Regulation of EGFR by SNPs

References

A Functional Common Polymorphism in a Sp1 Recognition Site of the Epidermal Growth Factor Receptor Gene Promoter

Wanqing Liu, Federico Innocenti, Michael H. Wu, et al.