Interethnic Difference in the Allelic Distribution of Human Epidermal Growth Factor Receptor Intron 1 Polymorphism

Wanqing Liu, Federico Innocenti, Peixian Chen, Soma Das, Edwin H. Cook, Jr., and Mark J. Ratain

ABSTRACT

Purpose: Epidermal growth factor receptor (EGFR) plays a critical role in signal transduction and is a target for a novel class of anticancer agents that aim to inhibit EGFR-mediated cancer cell growth. Previous studies have demonstrated a dinucleotide (CA)n repeat polymorphism in intron 1 of EGFR, ranging from 14 to 21 repeats, that has been suggested 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. Therefore, the evaluation of the allelic distribution of this polymorphism in populations of various ethnic origins will be crucial to understand the interindividual variability in EGFR expression.

Experiment Design: We evaluated the influence of ethnic- nicity on this polymorphism by genotyping individuals of Caucasian (n = 183), African-American (n = 84), and Asian (n = 66) background.

Results: The frequency of one of the longer alleles, allele 20 is significantly higher in Asian individuals (63% compared with 21% in Caucasians, P = 2 × 10^-18). In confirmation of prior studies, the shorter allele 16 was the most common allele in Caucasians (43%) and African-Americans (42%), but its frequency was significantly lower in Asians (average 17%, P = 10^-7 compared with Caucasians).

Conclusion: Major interethnic differences in the allelic frequencies of the EGFR intron 1 polymorphism exist. Our results may contribute to a better understanding of the molecular basis underlying ethnic differences in drug response and may be helpful for future strategies of individualized therapy with EGFR inhibitors.

INTRODUCTION

Human EGFR gene is located on chromosome 7p12.1–12.3. Its product, a Mᵦ 170,000 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 (reviewed in Ref. 1) and has been significantly associated with disease stage, prognosis, survival, and response to chemotherapy (2–11). Therefore, EGFR is an attractive target for treatment and prevention of cancer. Recent studies have demonstrated encouraging results with three EGFR-targeting inhibitors: cetuximab (IMC-225), a monoclonal antibody with affinity to the extracellular domain of EGFR, and two small EGFR-specific tyrosine kinase inhibitors, gefitinib (ZD1839) and erlotinib (OSI-774; Refs. 1, 12–16). Preliminary results in patients receiving gefitinib showed the presence of interethnic differences in drug response, with Japanese patients having a higher response rate than Caucasians (17).

A highly polymorphic (CA)n repeat was identified in intron 1 of the EGFR gene, and allelic frequencies in Caucasians were reported (18). Eight alleles from repeat numbers 14–21 were found. Allele 16 (alleles named according to the number of repeats) showed the highest frequency (42%), followed by allele 20 (26%), 18 (20%), and 21 (5%). Other alleles were relatively rare (18).

Recent in vitro and in vivo studies demonstrated that EGFR gene transcription activity declines with increasing number of CA repeats. The longer allele 21 showed 80% reduction of gene expression compared with the shorter allele 16 (19). In another study on breast cancer tissue, a similar effect of this (CA)n polymorphism was reported (20).

Because of the fact that limited data are available on the frequency distribution of this polymorphism in populations of different ethnicity, here we report the allelic distribution of the (CA)n polymorphism in Caucasian, African-American, and Asian populations.

MATERIALS AND METHODS

Samples. A total of 333 samples were analyzed, including 183 Caucasian, 84 African-American, and 66 Asian sam-

1 The abbreviation used is: EGFR, epidermal growth factor receptor.
Ethnic Differences in EGFR Intron 1 Allele Distribution

RESULTS

In 666 chromosomes, 9 alleles with 14–22 CA repeats and 2 rare alleles with 9 and 23 repeats were found (Table 1). The allele frequencies in Caucasians were in agreement with previous data (18). However, the frequency of allele 20 was much higher in Asian individuals (varying from 6 to 35% in different Asian groups, average 17%; $\chi^2 = 27.8; P = 10^{-7}$ compared with Caucasians). Concerning genotype frequencies, a genotype of 20/20 was the most common genotype in Asians with significantly higher frequency (average 41%) compared with Caucasians (7%; $\chi^2 = 41.11$; $P = 1.44 \times 10^{-10}$) or African-Americans (1%), whereas 16/16 was the most common genotype in both Caucasians (43%) and African-Americans (42%), but its frequency was significantly lower in Asians (varying from 6 to 35% in different Asian groups, average 17%; $\chi^2 = 27.8; P = 10^{-7}$ compared with Caucasians). Concerning genotype frequencies, a genotype of 20/20 was the most common genotype in Asians with significantly higher frequency (average 41%) compared with Caucasians (7%; $\chi^2 = 41.11$; $P = 1.44 \times 10^{-10}$) or African-Americans (1%), whereas 16/16 was the most common genotype in both Caucasians (43%) and African-Americans (42%) but not in Asians (average 5%; Fig. 1). In Asian populations, allele 16 showed much lower frequency in Chinese (6%,) compared with non-Chinese Asians (21%; $P = 0.0045$ compared to Non-Chinese Asian populations, Fisher’s Exact test, two-tailed).

Table 1 Allelic distribution of EGFR intron 1 (CA)\textsubscript{n} polymorphism in Caucasian, African-American, and Asian populations

<table>
<thead>
<tr>
<th>Allele</th>
<th>Caucasians</th>
<th>African-Americans</th>
<th>Chinese</th>
<th>Non-Chinese Asians</th>
<th>Total Asians</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>172</td>
<td>0.25</td>
<td>172</td>
<td>0.25</td>
<td>172</td>
</tr>
<tr>
<td>14</td>
<td>36</td>
<td>0.05</td>
<td>36</td>
<td>0.05</td>
<td>36</td>
</tr>
<tr>
<td>15</td>
<td>57</td>
<td>0.08</td>
<td>57</td>
<td>0.08</td>
<td>57</td>
</tr>
<tr>
<td>16</td>
<td>276</td>
<td>0.41</td>
<td>276</td>
<td>0.41</td>
<td>276</td>
</tr>
<tr>
<td>17</td>
<td>84</td>
<td>0.12</td>
<td>84</td>
<td>0.12</td>
<td>84</td>
</tr>
<tr>
<td>18</td>
<td>116</td>
<td>0.17</td>
<td>116</td>
<td>0.17</td>
<td>116</td>
</tr>
<tr>
<td>19</td>
<td>112</td>
<td>0.17</td>
<td>112</td>
<td>0.17</td>
<td>112</td>
</tr>
<tr>
<td>20</td>
<td>205</td>
<td>0.31</td>
<td>205</td>
<td>0.31</td>
<td>205</td>
</tr>
<tr>
<td>21</td>
<td>142</td>
<td>0.21</td>
<td>142</td>
<td>0.21</td>
<td>142</td>
</tr>
<tr>
<td>22</td>
<td>48</td>
<td>0.07</td>
<td>48</td>
<td>0.07</td>
<td>48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of chromosomes</th>
<th>366</th>
<th>168</th>
<th>52</th>
<th>80</th>
<th>132</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% confidence interval for frequency of allele 16</td>
<td>0.38–0.48</td>
<td>0.34–0.49</td>
<td>0.0079–0.12</td>
<td>0.15–0.35</td>
<td>0.11–0.24</td>
</tr>
</tbody>
</table>

* Freq. refers to allelic frequency and is expressed as percentage.

$^a$ Allele 20, $P = 0.0045$ compare to Non-Chinese Asian populations, Fisher’s Exact test, two-tailed.

$^b$ Allele 16, $\chi^2 = 27.8$, degrees of freedom = 1, $P = 10^{-7}$ compared with Caucasians.

$^c$ Allele 20, $\chi^2 = 76.8$, degrees of freedom = 1, $P = 2 \times 10^{-18}$ compared with Caucasians.

GENOTYPING AND DATA ANALYSIS

PCR was performed to amplify the target sequence. Primers were designed as described previously (18). One primer was labeled with fluorescein at the 5’ end. The genotyping procedure was performed according to a protocol reported previously (21). Fragment length was determined by direct sequencing. Capillary electrophoresis for sequencing and genotyping was performed on an ABI 3700 DNA sequencer in the DNA sequencing and genotyping core of the University of Chicago Cancer Research Center. Significance was performed either by Fisher’s exact or $\chi^2$ tests with $P < 0.05$ regarded as the cutoff for statistical significance.
DISCUSSION

Interethnic differences in the distribution of genetic polymorphisms in drug metabolizing enzymes, targets, receptors, and transporters are very common (22). For example, the variants in the thiopurine methyltransferase gene showed significant interethnic difference in global populations (22); similar findings were observed for a common length polymorphism in the thymidylate synthase 5’ untranslated region (23, 24). The population differences in the allelic distribution of EGFR intron 1 (CA)n polymorphism may be attributable to either the natural selection of an advantageous allele by unknown environmental factors or the fixation of allele frequency through a founder effect. Data from many genome-wide polymorphisms in Chinese compared with Caucasians and other populations indicated the contributions of a founder effect and population expansion to allelic population structure (22–25).

This population disparity is also an important factor that may contribute to interindividual variation in drug response and toxicity (22). Given the association between gene expression and the polymorphic (CA)n repeat, our study suggests that the response to anticancer therapies with EGFR inhibitors might vary in patients with different genotype backgrounds. Our data suggest that the higher response rate in the Japanese may be because of reduced EGFR expression associated with allele 20 (the most common allele in Asians; Ref. 17).

The most common adverse effect of EGFR inhibitors is cutaneous toxicity, such as skin rash (1). The appearance of skin rash seems to be associated with occurrence of antitumor response but is not clearly related to the dose level administered (26), also supporting the hypothesis that response to EGFR inhibitors has a genetic component. Given the high expression of EGFR in the epidermal tissue (27) and the proposed effect of (CA)n polymorphism on EGFR expression, it is plausible that the (CA)n genotype is associated with the sensitivity of both epidermal and tumor tissues to EGFR inhibitors, although other genetic determinants such as effectors in the signal transduction cascade may also contribute to this association.

Overexpression of EGFR in cancer cells has been significantly shown to be associated with disease stage, prognosis, survival, and response to chemotherapy (2–11), whereas the level of EGFR expression is primarily regulated by the abundance of its mRNA (28, 29). Therefore, the (CA)n polymorphism may at least partly contribute to the overexpression of EGFR (20) and influence the clinical outcome. In a preliminary report of an association study between this polymorphism and clinical response to 5-fluorouracil/oxaliplatin treatment in patients with metastatic colorectal cancer, patients with genotype 16/16 showed poor drug response and overall survival compared with those with 16/18 and 16/20 genotypes (30).

Our study provides an important starting point for better understanding the molecular basis underlying ethnic differences in response to EGFR inhibitors and may be helpful for the future strategies of individualized therapy with molecules targeting EGFR. However, there are some limitations in our study, and several questions still should be addressed. For example, the sample size of Asian samples is relatively small. It should also be emphasized that other polymorphisms in linkage disequilibrium with the (CA)n repeat may also contribute to variability in EGFR expression. Therefore, additional studies with large sample size are required to evaluate the haplotype, relationship among genotypes, EGFR expression, and clinical outcomes of EGFR-targeting drugs.

REFERENCES

Interethnic Difference in the Allelic Distribution of Human Epidermal Growth Factor Receptor Intron 1 Polymorphism

Wanqing Liu, Federico Innocenti, Peixian Chen, et al.