Ethnic variation in genotype frequencies of a p53 intron 7 polymorphism

Petra Berggren1,2,4, Rajiv Kumar1, Gunnar Steineck2, Masayoshi Ichiba3 and Kari Hemminki1

1Department of Biosciences at Novum, Karolinska Institute, 141 57 Huddinge, Sweden, 2Clinical Cancer Epidemiology and Clinical Oncology, Department of Oncology-Pathology, Karolinska Hospital, 171 76 Stockholm, Sweden and 3Department of Community Health Science, Saga Medical School, Saga 849, Japan

Using a PCR–restriction enzyme-based method we found large ethnic variations when a C→T polymorphism in the human p53 gene at position 14181 in intron 7 was studied in Finnish, Polish, Hungarian, Italian, Japanese, Indian and Chinese populations. The largest variations were found between Caucasian and Asian, genotype frequencies varied from C 0.95 and T 0.05 in the Finnish population to C 0.67 and T 0.33 in the Chinese population, the relative risk (RR) for T being 6.5 (95% CI 3.4–12.3, P < 0.001). Variations were also found between Finnish and Italian and between Italian and Chinese (RR for T 2.4, 95% CI 1.7–4.2, respectively).

Introduction

The human tumour suppressor gene p53 has been highly conserved during evolution; only two non-synonymous and three synonymous polymorphisms in the coding regions and eight polymorphisms in the intronic regions have been found (Hernandez-Boussard et al., 1999). Ethnic variations in geno-type frequencies of different polymorphisms are common in several genes, such as the human drug metabolizing enzymes cytochrome P-450 (CYP450), glutathione S-transferase (GSTM1) and N-acetyltransferase (NAT2), but there are also reported variations in genotype frequencies of the p53 codon 72 polymorphism (Arg/Pro), that range from 0.79/0.21 in the Mediterranean population to 0.47/0.53 in the African-American population (Jin et al., 1995; To-Figueras et al., 1996).

We have previously studied a C→T polymorphism located at position 14181, 72 bp downstream of the 3’-end of exon 7 of the human p53 gene (GenBank accession no. X54156) that is in complete linkage disequilibrium with a T→G polymorphism 20 bp further downstream at position 14201. The study was done in urinary bladder cancer patients and non-cancer controls of Caucasian origin and allele frequencies were C 0.92 and T 0.08 in both groups (Berggren et al., 2000). In this study we have investigated these linked polymorphisms in several ethnic groups to establish variations in genotype frequencies.

Materials and methods

A total of 485 samples from seven different populations were analysed. DNA was extracted from normal tissue surrounding the tumour in 44 Hungarian lung cancer patients and 22 Chinese oesophageal cancer patients. Normal tissue was not available from two of the Chinese oesophageal cancer patients, so tumour tissue was used.

We used a PCR–restriction enzyme treatment method previously described in detail (Berggren et al., 2000). Briefly, PCR was performed for 32 cycles in a total volume of 10 µl containing 10 ng genomic DNA and the following primers: sense, 5’-GTT GCC TCT GAC TGT ACC ACC-3’; antisense, 5’-GGC GGA AAT GTG ATG AGA-3’. After PCR samples were treated with the restriction enzyme EcoRI (an isoschizomer of Avall) that specifically cleaves at 5’-GAC/ACC-3’. Thus only samples heterozygous for CT or homozygous for TT at 14181 were cut, leaving two fragments of 180 and 61 bp for the TT homozygous samples and an additional 241 bp fragment from the C allele for CT heterozygous samples. Polymorphisms were easily detected in a 5% PAGE gel stained with ethidium bromide (see Figure 1). All samples that were heterozygous CT at position 14181 were re-amplified by a second PCR and treated with the restriction enzyme Bsp120I (a neoschizomer of Apal) that specifically cleaves at 5’-GAC/CCC-3’. Thus only samples that were CC homozygous or CT heterozygous at 14181 were cut, leaving two fragments of 179 and 62 bp. Bsp120I was used to exclude the possibility that the CT heterozygous samples were the result of a partial EcoRI cut of TT homozygous samples. We found no samples that were partially cut with EcoRI, indicating that all detected heterozygous samples were true heterozygotes.

Genotype frequencies were calculated from observed genotypes and then compared with the Finnish population as a reference group (lowest allele frequency for T). Expected genotype frequencies were calculated from the Hardy–Weinberg equilibrium (HWE): \( p^2 + 2pq + q^2 = 1 \), where \( p \) and \( q \) were the observed genotype frequencies (\( p + q = 1 \)). Observed and expected genotypes were compared with a Pearson \( \chi^2 \) test as follows:

\[
Q = \frac{(\text{observed CC} – \text{expected CC})^2}{\text{expected CC}} + \frac{(\text{observed CT} – \text{expected CT})^2}{\text{expected CT}} + \frac{(\text{observed TT} – \text{expected TT})^2}{\text{expected TT}} \]

using \( \chi^2 = 3.84, df = 1 \) [three observations, one estimated parameter (\( p \), \( q \), \( Q \)). 0.05 (Microsoft Excel 97) (Hartl and Clark, 1997). The Mantel–Haenszel method gave the 95% confidence interval for RR and Fisher’s exact test was used to obtain P.

Results and discussion

Large differences were found between the genotypes C and T at position 14181 of the p53 gene among the different populations (see Table I). Significant differences in genotype frequencies were found between the Finnish population, compared with the Italian, Indian, Japanese and Chinese populations; and also between the Italian, Japanese and Chinese populations, but not between the Italian and Indian populations. Relative risk (RR), confidence intervals (CI) and P values are shown in Table I. Three samples failed to amplify repeatedly (one Finnish, one Polish and one Japanese). These samples were also tested with primers for other genes located on chromosomes other than 17. No amplification was seen for the Japanese sample with any other marker and the two other samples only amplified as short fragments (100–150 bp), whereas longer fragments failed to amplify (>200 bp), suggesting poor quality DNA. These three samples were excluded from the calculations.

All genotype frequencies corresponded to the HWE except for two groups of Chinese samples that, for unknown reasons, showed significant deviation from the HWE (see Table I). Also included in Table I is a Swedish population from a previous study of urinary bladder cancer patients and controls.
No difference was seen between the two groups and they were treated as one group (Berggren et al., 2000).

Previous findings show that there is complete linkage between C at 14181 and T at 14201 and between T at 14181 and G at 14201 (Prosser and Condie, 1991; Berggren et al., 2000). As a test to determine whether this was also true for these samples, tumour tissues from the Chinese oesophageal cancer patients (n = 22) were sequenced using previously reported methods (Berggren et al., 2000). All Chinese oesophageal cancer samples that were sequenced showed complete linkage. Samples that were heterozygous CT (14181) were also heterozygous TG (14201). A representative sequencing picture has been published previously (Berggren et al., 2000). Four heterozygous samples showed loss of heterozygosity (LOH): two of these had lost the C(14181) T(14201) allele and two had lost the T(14181) G(14201) allele. As previously shown, complete linkage disequilibrium was found between the two polymorphisms.

Significant linkage disequilibrium is often found for genes and polymorphisms that are tightly linked (i.e. located near each other on the same chromosome). Since there is always an element of chance in the genetic make-up of the two gametes that will form a new generation, chance alone can result in changes in allele frequency, a process known as random genetic drift. Due to random genetic drift each allele faces a risk of extinction in each generation and therefore a single period of small population size, called a population bottleneck, can result in a reduced number of alleles (Hartl and Clark, 1997). The linked polymorphisms we have studied may have resulted from two separate mutations where the original non-mutated allele has disappeared through a population bottleneck. It is also possible that the linked polymorphisms are the result of two simultaneous mutations. Linked polymorphisms can be brought about by recombination, however, since the studied polymorphisms are located only 20 bp apart, recombination between them is extremely rare. Another explanation for linkage disequilibrium is natural selection favouring some genotypes over others, however, because the polymorphisms are located in the intronic sequence this is not likely as an explanation, but cannot be ruled out without functional studies of the protein.

In this study of nearly 500 samples we have investigated genotype frequencies for the C→T polymorphism at position 14181, 72 bp downstream of the 3’-end of exon 7 of the human p53 gene, in different ethnic groups. We have shown that genotypes vary 6-fold (proportion with any T, either one, i.e. CT, or two, i.e. TT) for the p53 intron 7 polymorphism between different populations. The variations seem to be largest between groups from different continents, for example Caucasian (Finnish and Swedish) and Asian (Indian, Japanese and Chinese), but there are also differences between Finnish and Italians. Since the genotypes CC, CT and TT (at 14181) do not occur at the same frequency within the different populations, ideally a much larger number of samples than we have analysed should be used to obtain absolute statistical validity for the study. Given the expected frequencies of TT

![Fig. 1. 5% PAGE gel showing, from the left: lanes 1–3, treated with the restriction enzyme Eco47I (CC, two intact alleles; CT, one cut allele (T) and one intact allele (C); TT two cut alleles); lanes 4–6, treated with Bsp120I (CC, two cut alleles; CT, one cut allele (C) and one intact allele (T); TT, two intact alleles); lane 7, a negative control (;).](image)

**Table 1. Variation in genotype frequencies for p53 intron 7 C→T polymorphism**

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>C</th>
<th>T</th>
<th>RR</th>
<th>CI</th>
<th>P</th>
<th>RR</th>
<th>CI</th>
<th>P</th>
<th>Q</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>99&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.949</td>
<td>0.051</td>
<td>1.0</td>
<td>Ref.</td>
<td></td>
<td>0.33</td>
<td></td>
<td>0.56</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>Sweden&lt;sup&gt;b&lt;/sup&gt;</td>
<td>330</td>
<td>0.923</td>
<td>0.077</td>
<td>1.4</td>
<td>0.7–2.8</td>
<td></td>
<td>0.30</td>
<td></td>
<td>0.55</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>89</td>
<td>0.927</td>
<td>0.073</td>
<td>1.6</td>
<td>0.7–3.4</td>
<td></td>
<td>0.20</td>
<td></td>
<td>0.44</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>44&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.909</td>
<td>0.091</td>
<td>1.8</td>
<td>0.7–4.4</td>
<td></td>
<td>0.01</td>
<td>1.0</td>
<td>Ref.</td>
<td>0.23</td>
<td>0.63</td>
</tr>
<tr>
<td>Italy</td>
<td>103</td>
<td>0.879</td>
<td>0.121</td>
<td>2.4</td>
<td>1.2–4.9</td>
<td></td>
<td>0.01</td>
<td>&lt;0.001</td>
<td>1.9</td>
<td>1.0–3.5</td>
<td>0.074</td>
</tr>
<tr>
<td>India</td>
<td>26&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.769</td>
<td>0.231</td>
<td>4.6</td>
<td>2.1–10.0</td>
<td></td>
<td>&lt;0.001</td>
<td>2.4</td>
<td>1.5–3.9</td>
<td>&lt;0.001</td>
<td>2.10</td>
</tr>
<tr>
<td>Japan</td>
<td>46</td>
<td>0.707</td>
<td>0.293</td>
<td>5.8</td>
<td>2.9–11.5</td>
<td></td>
<td>&lt;0.001</td>
<td>2.6</td>
<td>1.6–4.2</td>
<td>&lt;0.001</td>
<td>3.90</td>
</tr>
<tr>
<td>China</td>
<td>37&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.689</td>
<td>0.311</td>
<td>6.2</td>
<td>3.1–12.3</td>
<td></td>
<td>&lt;0.001</td>
<td>2.8</td>
<td>1.5–5.2</td>
<td>&lt;0.001</td>
<td>0.97</td>
</tr>
<tr>
<td>China&lt;sup&gt;f&lt;/sup&gt;</td>
<td>16</td>
<td>0.656</td>
<td>0.344</td>
<td>6.8</td>
<td>3.1–14.7</td>
<td></td>
<td>&lt;0.001</td>
<td>2.8</td>
<td>1.6–4.9</td>
<td>0.001</td>
<td>5.89</td>
</tr>
<tr>
<td>China&lt;sup&gt;g&lt;/sup&gt;</td>
<td>22&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.659</td>
<td>0.341</td>
<td>6.8</td>
<td>3.3–14.0</td>
<td></td>
<td>&lt;0.001</td>
<td>2.8</td>
<td>1.6–4.9</td>
<td>0.001</td>
<td>9.93</td>
</tr>
<tr>
<td>China all</td>
<td>75&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.673</td>
<td>0.327</td>
<td>6.5</td>
<td>3.4–12.3</td>
<td></td>
<td>&lt;0.001</td>
<td>2.7</td>
<td>1.7–4.2</td>
<td>&lt;0.001</td>
<td>9.93</td>
</tr>
</tbody>
</table>

n, number of samples; C = cytosine, T = thymidine at position 14181 of the p53 gene. The Mantel–Haenszel equation was used to calculate RR<sup>1,2</sup> (relative risk) and 95% CI (95% confidence interval) and Fisher’s exact test was used to obtain the P value. Two comparisons are made; the first uses the Finnish population as the reference group and the second uses the Italian population as the reference group. Observed and expected genotype frequencies were compared using the χ² test using χ² < 3.84, df 1, α 0.05. Test results are represented by Q.

<sup>a</sup>Normal tissue from cancer patients.

<sup>b</sup>Previously published data (Berggren et al., 2000).

<sup>c</sup>Healthy Chinese.

<sup>d</sup>Chinese with cancer.

<sup>e</sup>All Chinese.
homozygous samples (the smallest frequencies), as a statistical rule of thumb a large enough sample size should be used so that the expected frequency of TT homozygotes (p) multiplied by the total number of patients (n) should give at least five (n×p) expected TT homozygous individuals (Blom, 1989).

If the polymorphism is associated with a risk of cancer, the genotype frequency in cancer patients will be different from that of the general population. Our data suggests that this is not the case for urinary bladder cancer (Berggren et al., 2000) and we are not aware of any data indicating that this polymorphism elevates cancer risk. The Polish, Italian, Japanese and Chinese (n = 53) samples studied were from healthy individuals and showed a similar variation in polymorphism distribution between populations as did the cancer patients. Thus all available data suggests that our results are valid for the general population.

Many reports on ethnic differences in genotype frequencies of polymorphisms have considered the enzymes involved in the activation/inactivation of known carcinogens, such as the CYP450, GSTM1 and NAT2 genes (d’Errico et al., 1996; Smith et al., 1998). These are especially important when they influence mutations in key regulators of cell cycle control, such as the tumour suppressor gene p53 (Ryberg et al., 1994). Individuals who inherit susceptibility alleles of CYP1A1 and GSTM1 have been found to have a high frequency of p53 mutations as a consequence of carcinogen exposure (Kawajiri et al., 1996). The CYP1A1 T5639C polymorphism flanking the 3'-region has only been found in African-American and African populations (Crofts et al., 1993) and the T6235C (flanking the 3'-region) and A4889G (in exon 7 codon 462) polymorphisms are closely, but not completely, linked in Caucasian and Asian (Hayashi et al., 1991; Hirvenon et al., 1992). (Inoue et al., 2000) have studied polymorphisms in CYP1A1 and CYP1B1. They found race-related differences in the occurrence of polymorphisms in both genes and suggested that this may, in part, cause differences in the incidence of lung and breast cancer in Japanese and Caucasian populations. Linked polymorphisms have also been observed in tumour suppressor genes such as CDKN2A (Zhang et al., 1994; Aitken et al., 1999). Linked polymorphisms in exon three of the CDKN2A genes (p16 and p14ARF) have been associated with disease progression in malignant melanoma (Saurio et al., 2000). Ethnic variations have also been found in the p53 codon 72 Arg/Pro polymorphism (CGC/CCC), where allele frequencies varied from 0.790/0.21 in Mediterranean Caucasians to 0.47/0.53 in African-Americans (Jin et al., 1995; To-Figuera et al., 1996). Allele frequencies in Caucasians (English and Swedish) were 0.78/0.22 and 0.72/0.28 and were all close to 0.60/0.40 in Asians (Chinese and Japanese) (Kawajiri et al., 1993; Birgard et al., 1995; Minaguchi et al., 1998; Rosenthal et al., 1998; Yamashita et al., 1999; Lee et al., 2000).

We have detected population variation of a C→T polymorphism in the p53 gene. We did not find any evidence that the studied polymorphism elevates cancer risk. However, studies of polymorphisms may be important because genotype frequencies vary between populations and could possibly explain, in part, the differences in cancer incidence between populations. Intronic polymorphisms probably play a major role in the field and perhaps the most important polymorphisms are those changing protein function of metabolic enzymes and of cell cycle regulators.

Acknowledgements
The authors wish to thank Charlotte Granström for help with statistical calculations and Drs Sai-Mei Hou and Kamila Pina for providing samples from healthy Chinese.

References
P.Berggren et al.


Received on March 8, 2001 and accepted on June 28, 2001