Electronic Microarray Technique for Detection of Nine Base Substitutions Including Single-Nucleotide Polymorphisms in the Human OGG1 Gene, Zhilong Gong,1,2 Christine Teixeira,3,4 James Z. Xing,1 Jun Yokota,2 Takashi Kohno,2 Stephan Gabos,2 X. Chris Le,5 and Xing-Fang Li5
(1) Environmental Health Sciences, Department of Public Health Sciences, University of Alberta, Edmonton, Alberta, Canada; (2) Biology Division, National Cancer Center Research Institute, Tokyo, Japan; (3) Health Surveillance Branch, Population Health Division, Alberta Health and Wellness, Edmonton, Alberta, Canada; # the first two authors contributed equally to this work; * address correspondence to these authors at: Department of Public Health Sciences, University of Alberta, Edmonton, Alberta T6G 2C3, Canada; fax 780-492-7800, e-mail x.c.le@ualberta.ca or xingfang.li@ualberta.ca)

Single-nucleotide polymorphisms (SNPs) are potentially important biomarkers for molecular diagnostics. Several SNPs and somatic mutations have been characterized in a human 8-hydroxyguanine glycosylase gene, OGG1, which codes for a DNA repair enzyme responsible for excising 8-hydroxyguanine from damaged DNA (1, 2). The distribution of genotypes for a SNP, S326C, in the OGG1 gene has been implicated in the susceptibility to various cancers, including lung, esophageal, and prostate cancer (2–4). A functional consequence of this SNP is a difference in the activity to suppress mutagenesis induced by 8-hydroxyguanine. Another SNP, R46Q, causes alternative splicing as well as a difference in the mutation suppressive activity. In addition, somatic mutations in the OGG1 gene have been detected in a subset of lung, gastric, and kidney cancers, suggesting that aberrations of OGG1 contribute to human carcinogenesis by enhancing mutagenesis induced by 8-hydroxyguanine (2–5).

The objective of this study was to develop a novel technique that can provide multiplex analysis of base substitutions including SNPs in the OGG1 gene by use of electronic microarray technology. Here we report the development of an electronic microarray technique for analysis of all nine base substitutions in 1-kb amplicons. The technique was successfully applied to the analysis of the four SNPs and five mutations in the OGG1 gene in four cancer cell lines with a double-blind experimental design, and the results were in excellent agreement with the previous identification of SNPs in the cell lines by DNA sequencing and single-strand conformation polymorphism gel electrophoresis. For quality control, we used site-directed mutagenesis and megaprimer PCR techniques (6, 7) to develop a 1-kb DNA consisting of all base substitutions in the OGG1 gene. The present technique has the unique capability of simultaneous detection of multiple base substitutions with internal validation for the analysis of each substitution. This is the first report to demonstrate the electronic transport of 1-kb DNA samples to microarray spots.

The assay involves three steps, immobilization of capture oligonucleotides, hybridization with targets, and detection of reporter fluorescence; these steps are shown in more detail in Fig. 1 of the Data Supplement that accompanies the online version of this Technical Brief at http://www.clinchem.org/content/vol50/issue8/. A capture oligonucleotide was designed to contain the sequence located immediately downstream of a base substitution in the OGG1 gene. We designed and optimized nine capture oligonucleotides for the nine base substitutions in OGG1. To optimize a capture oligonucleotide, we examined numerous oligonucleotides containing 20–35 nucleotides by adding or subtracting one base at a time and obtaining their melting temperatures and secondary structures by use of a DNA melting folder provided by The Zuker Group (www.bioinfo.rpi.edu/applications/mfold/) and the “Oligo Analyzer 2.5” software (www.IDTDNA.com) provided by Integrated DNA Technologies. The optimized capture oligonucleotides that had a melting temperature (Tm) between 58.5 and 70.4°C with minimal amount of secondary structure were chosen (Table 1).
We designed a pair of reporters to differentiate between the wild type and mutants of a specific allele. The reporters generally consisted of 9–12 nucleotides, and their sequences were complementary to the corresponding targets, which are located upstream of the SNP including the substitution site (adjacent to the capture). Several factors were considered when we designed the reporters to obtain the best differentiation between the wild type (major allele for SNP) and mutants (minor allele for SNP). These include $T_m$, which was optimized for the targeted SNPs and minimized for the nontargets; base-stacking energy; GC content; and secondary structure of the reporters. The differences between the $T_m$ of the wild-type and mutant reporters were generally $\pm 3°C$, except for the A85S reporters, which differed by 7.4°C, and the K249Q reporters, which differed by 5.6°C. The larger differences in the $T_m$ of the A85S reporters are attributable to unbalanced base-stacking energy and high GC content. The base-stacking energies for the wild-type and mutant reporters of A85S are $-14.6$ and $-9.81$ kcal/mol, respectively. As for K249Q, the reporters were $\pm 8$ bases upstream of the SNP and had GC content of 75% for the wild-type and 87.5% for the mutant alleles. The best suitable reporters were chosen with a $T_m$ of 26.5–43.1°C with minimum or no secondary structure. Reporters specific for the wild-type and mutant alleles were labeled at the 5’ end with Cy3 (green) and Cy5 (red), respectively. After series testing using mutant and wild-type DNA, we obtained nine pairs of reporters (Table 1) that provided the best ratios of the red and green fluorescence intensities to differentiate between the wild-type and mutant alleles for each of the nine base substitutions in the OGG1 gene.

We first analyzed a single SNP, S326C, in four human cancer cell lines. S326C represents a change of a single nucleotide (from C to G) at codon 326, which changes an amino acid from Ser to Cys in the OGG1 protein. Our analysis correctly identified the single nucleotide difference between the S326 and C326 alleles in all four cell lines (see Fig. 2 in the online Data Supplement).

To further improve the specificity, we optimized the temperatures for thermal stringency (see Fig. 3 in the online Data Supplement for an example of the effect of temperature on the analysis of the SNP S326C). They were 42–44°C for R46Q, 33–35°C for A85S, 37–39°C for R131Q, 38–40°C for R154H, 33–35°C for S232T, 41–43°C for K249Q, 31–33°C for R304W, 30–32°C for G308E, and 31–33°C for S326C.

To facilitate quality control of fluorescence measurements and quality assurance of the identification of wild-type and mutant alleles, we developed two 1-kb oligonucleotide DNAs for the nine base substitutions in OGG1. PCR-based site-directed mutagenesis was used to introduce nine mutations into a cDNA fragment encoding the OGG1 transcript variant 1a (codons 12–335). The procedures for generating the mutant OGG1 DNA fragments are described in Fig. 4 of the online Data Supplement. Human OGG1 cDNA sequences were obtained from GenBank (www.ncbi.nlm.nih.gov), and this cDNA was kindly provided by Dr. G.L. Verdone (Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA). The forward and reverse PCR primers consisted of 5’-GGGCATCGTACTCTAGCCTCC-3’ and 5’-TGCTGTGGCGTCCTCTGAGCAATGCCG-3’, respectively. DNA sequencing analysis confirmed incorporation of the nine mutations in a full-length mutant DNA. A mixture of the wild-type and the mutant DNAs at a ratio of 1:1 was used as a heterozygous control, and the fluorescence signals were normalized to 1. The use of these internal standards eliminated the effect of the background signal and other potential variations attributable to nonuniform

<table>
<thead>
<tr>
<th>SNP</th>
<th>Reporter</th>
<th>$T_m$ °C</th>
<th>Capture*</th>
<th>$T_m$ °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>R46Q wild-type</td>
<td>5’-Cy3/GAC AAT TTC G-3’</td>
<td>37.1</td>
<td>5’-GTG GAG GGA GCA AAG TCC TGCC ACG TGG AGC- /3Bio/</td>
<td>69</td>
</tr>
<tr>
<td>R46Q mutant</td>
<td>5’-Cy5/GAC AAT TTC A-3’</td>
<td>35.1</td>
<td>5’-CTA GCA GGC CCA CAC CAG ACG AGC TGAG GAG- /3Bio/</td>
<td>70.4</td>
</tr>
<tr>
<td>A85S wild-type</td>
<td>5’-Cy3/AGG GCCAAG-3’</td>
<td>27.1</td>
<td>5’-ACT GCT GGC ACA AGA CCC CAT CC/ /3Bio/</td>
<td>65.1</td>
</tr>
<tr>
<td>A85S mutant</td>
<td>5’-Cy5/CAA GAG CCA GT-3’</td>
<td>34.5</td>
<td>5’-GGG CGA TGT TGT TGT AGG AGG AAC AGC- /3Bio/</td>
<td>63.8</td>
</tr>
<tr>
<td>R131Q wild-type</td>
<td>5’-Cy3/AGG TGT GGC-3’</td>
<td>30.7</td>
<td>5’-CAT ATG AGG CCC ACA AGG CCC- /3Bio/</td>
<td>65.1</td>
</tr>
<tr>
<td>R131Q mutant</td>
<td>5’-Cy5/AGG TGT GGC A-3’</td>
<td>31.4</td>
<td>5’-AGG TGG TCT GCT GCA TCT GCC TGA TGG/ /3Bio/</td>
<td>68.0</td>
</tr>
<tr>
<td>R154H wild-type</td>
<td>5’-Cy3/GCC AGT GTG GC-3’</td>
<td>38.0</td>
<td>5’-GCC TGT GCC CTG TGCC TGA AAG GAC- /3Bio/</td>
<td>69.5</td>
</tr>
<tr>
<td>R154H mutant</td>
<td>5’-Cy5/GCC AGT GTG A-3’</td>
<td>35.2</td>
<td>5’-ACC ATC TGC TGG CTC GGC ACG- /3Bio/</td>
<td>58.5</td>
</tr>
<tr>
<td>S232T wild-type</td>
<td>5’-Cy3/CGA GAG TCC T-3’</td>
<td>29.7</td>
<td>5’-ATT GGC GCA GGT CGC CAC ACG- /3Bio/</td>
<td>65.6</td>
</tr>
<tr>
<td>S232T mutant</td>
<td>5’-Cy5/CGA GAG TCC A-3’</td>
<td>30.3</td>
<td>5’-ATT GGC GCA GGT CGC CAC ACG- /3Bio/</td>
<td>65.6</td>
</tr>
</tbody>
</table>

* Bio, Biotin.
hybridization, difference in laser intensities, and differences in the fluorescence yields of the two fluorophores. Having established an assay with quality control, we analyzed four cancer cell lines for all nine base substitutions: R46Q, A85S, R131Q, R154H, S232T, K249Q, R304W, G308E, and S326C. The H1299 lung cancer cell line was
provided by Drs. J.D. Minna and A.F. Gazdar of
the University of Texas Southwestern Medical Center (Dallas, TX). Two other lung cancer cell lines, A427 and H526, and one gastric cancer cell line, MKN1, were obtained from the American Type Culture Collection and the Japan Cell Resource Bank and Institute for Fermentation, Japan. The total cellular RNA was isolated by use of Trizol (Invitrogen Life Technologies) and was reverse-transcribed with Superscript Reverse Transcriptase (Invitrogen). The sample preparation and analysis were carried out in a double-blind fashion. The cell samples were labeled with a number and given to a researcher to extract RNA and carry out reverse transcription-PCR without previous knowledge of the cell lines. Another researcher carried out the analysis of the nine base substitutions in these samples, using the microarray technique. The genotypes of the four cell lines were identified based on the ratios of fluorescence signals (Fig. 1). Cell line A427 was homozygous for the wild-type allele at the other eight sites (green: red ratio = 148), and homozygous for the wild-type allele (major allele) at the other eight sites (green: red ratio > 8). Cell line H526 was homozygous for Q46 and homozygous for the wild-type allele at the other eight sites. Cell line H1299 was homozygous for E308 and homozygous for the wild-type allele at the other eight sites. Cell line MKN1 was heterozygous for R154H and S326C and homozygous for the wild-type allele at the other seven sites. These results were in 100% agreement with genotypes determined by DNA sequencing and single-strand conformation polymorphism gel electrophoresis (8, 9).

We also examined the number of times that a capture-loaded microchip can be used. A microchip that had previously been used three times was incubated with 0.1 mol/L NaOH and washed with water three times to strip off the bound targets and reporters. The microchip was repeatedly used to analyze SNP G308E. After each analysis, the incubation with 0.1 mol/L NaOH and the series of three washes were repeated. Over a period of 3 days, four repeat analyses of the DNA fragment for SNP G308E showed that the signal ratios were in the range 5–79 (green:red) for G308 and 6–13 (red:green) for E308, demonstrating positive identification of the genotypes. Additional repeated use of the microchip let to signal-to-background ratios or red:green fluorescence intensity ratios < 5, the criteria for reliable identification of the alleles. These results suggest that the capture-loaded microchips can be reused for seven repeated analyses.

Unlike previous assays that use genomic DNA to measure allele distribution and frequency, we used mRNA as our source of genetic material. The advantages of screening amplified mRNA for polymorphisms include the use of a single reverse transcription-PCR reaction to generate a templates containing multiple SNPs and the potential to quantitatively assess the expression of minor alleles. Interestingly, OGG1 is also expressed as multiple transcripts, several of which were amplified by the primer pair we designed (see Fig. 5 in the online Data Supplement). Furthermore, accurate identification of base substitutions, including SNPs, was possible without separation of the variants.

The assay can be used to analyze a single base substitution in >30 samples (triplicate analyses for each) on a single chip. Alternatively, it can be used to simultaneously analyze all nine base substitutions in multiple samples. On average, analysis of each SNP takes 4 min and costs approximately US $1.70 for analysis of multiple samples.

In conclusion, we have shown that the electronic microarray is an excellent platform for accurate identification and high-throughput analysis of base substitutions, including SNPs. The use of a heterozygous control containing equal mixtures of wild-type and mutant DNA within a 1-kb fragment provides internal validation for SNP genotyping. The technique demonstrated robust detection of a single genetic variation by use of model oligonucleotides, small exons, or short fragments of DNA. The technique is potentially useful for association studies of SNPs in the OGG1 gene with susceptibility to diseases attributable to environmental exposure or for studies of the efficacy of radiation treatment of cancers related to the OGG1 SNPs.

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, Alberta Health and Wellness, the National Institutes of Health, the Canada Research Chair Program (X.C.L.), and the NSERC University Faculty Award (X.F.L.).

References

DOI: 10.1373/clinchem.2004.034009