Clinical Chemistry 46:10
1555–1561 (2000)

Molecular Diagnostics and Genetics

Evaluation of the Performance of a p53 Sequencing Microarray Chip Using 140 Previously Sequenced Bladder Tumor Samples

Friedrik P. Wikman,[1] Ming-Lan Lu,[2] Thomas Thykjaer,[1] Sanne H. Olesen,[1] Lars D. Andersen,[1] Carlos Cordon-Cardo,[2] and Torben F. Ørntoft[1*]

Background: **Testing for mutations of the *TP53* gene in tumors is a valuable predictor for disease outcome in certain cancers, but the time and cost of conventional sequencing limit its use. The present study compares traditional sequencing with the much faster microarray sequencing on a commercially available chip and describes a method to increase the specificity of the chip.**
Methods: **DNA from 140 human bladder tumors was extracted and subjected to a multiplex-PCR before loading onto the p53 GeneChip from Affymetrix. The same samples were previously sequenced by manual dideoxy sequencing. In addition, two cell lines with two different homozygous mutations at the *TP53* gene locus were analyzed.**
Results: **Of 1464 gene chip positions, each of which corresponded to an analyzed nucleotide in the sequence, 251 had background signals that were not attributable to mutations, causing the specificity of mutation calling without mathematical correction to be low. This problem was solved by regarding each chip position as a separate entity with its own noise and threshold characteristics. The use of background plus 2 SD as the cutoff improved the specificity from 0.34 to 0.86 at the cost of a reduced sensitivity, from 0.92 to 0.84, leading to a much better concordance (92%) with results obtained by traditional sequencing. The chip method detected as little as 1% mutated DNA.**
Conclusions: **Microarray-based sequencing is a novel option to assess *TP53* mutations, representing a fast and inexpensive method compared with conventional sequencing.**
© 2000 American Association for Clinical Chemistry

Mutations of the *TP53* gene contribute to tumor progression and are likely to provide relevant prognostic information to assist in the management of cancer patients. Several studies have reported that *TP53* gene mutations are associated with overall survival in different epithelial cancers *(1, 2)*. Introduction of a reliable high-throughput sequencing method into clinical laboratories would enable clinical decisions based on the identification of inactivating *TP53* mutations. A DNA microarray-based method is a promising alternative assay for such routine use because it is less time-consuming and of relatively low cost compared with conventional sequencing. Furthermore, the latter method has the drawback that acrylamide is a potentially hazardous constituent. Previously, DNA array-based sequencing has been applied to the *BRCA1* and the cystic fibrosis transmembrane conductance regulator genes, showing promising results on relatively small numbers of cases *(3, 4)*. In a recent study, a p53 oligonucleotide array (Affymetrix p53 GeneChip®) was compared with three other DNA-based mutation detection methods on 100 primary lung cancers and showed a high detection rate of missense mutations and a good sensitivity *(5)*.

The DNA array hybridization method is a technique based on the use of nonporous supports such as glass for immobilization and the development of confocal laser scanning. Lately, the adaptation of techniques developed by the computer industry to the manufacture of the arrays has sparked implementation of in situ synthesized high-density DNA microarrays. This photolithographic technique allows the synthesis of defined oligomers on 50-μm squares, and squares as small as 24 μm, so that a 1.28 × 1.28 cm chip can contain 65 536 50-μm squares, or 262 144 24-μm squares, each of which is covered with a different oligonucleotide *(6)*. Other approaches use presynthesized

[1] Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, 8200 Aarhus N, Denmark.
[2] Department of Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10021.
*Address correspondence to this author at: Department of Clinical Biochemistry, Skejby Sygehus, Brendstrupgaardsvej, 8200 Aarhus N, Denmark. Fax 45-89-49-60-18; e-mail orntoft@kba.sks.au.dk.
Received April 3, 2000; accepted July 7, 2000.

oligomers spotted on solid supports, membranes, or glass, techniques that allow a density of up to 3640 spots/cm^2. The first generation of p53 sequencing chips, Affymetrix p53 GeneChip, which is the object for this study, showed probe-specific background signals when applied to 140 bladder tumors. When results were corrected for these signals, a good correlation in the detection of missense mutations between the chip and conventional sequencing was observed. However, base pair insertions and deletions could not be detected. A dilution experiment showed that the chip was able to detect mutated DNA when it constituted <2% of total DNA, possibly making the GeneChip the most sensitive sequencing method known.

Materials and Methods

DNA was extracted from microdissected tissue sections from bladder cancer tumors collected at Memorial Sloan-Kettering Cancer Center, New York *(1)*. Two *TP53* mutated cell lines were kindly supplied by L. Noerum, University Hospital of Aarhus, Aarhus Kommunehospital, Denmark. The BT20 cell line contains a AAG→CAG mutation at codon 132, whereas the PancI cells harbor a CGT→CAT mutation at codon 273. The GeneChip p53 chips, wash station, p53 primer set, control DNA, fragmentation reagent, control oligonucleotide, and GeneChip software were from Affymetrix. Scanning of the chips was conducted with a HP GeneArray™ Scanner from Hewlett Packard. AmpliTaq Gold was from Perkin-Elmer, dNTPs were from Pharmacia Biotech, and fluorescein-N6-dideoxy-ATP was from DuPharma. PCR was performed on a PTC-200 DNA Engine. Additional data analysis was done with Quattro Pro 8 from Corel®. Acetylated bovine serum albumin and calf intestine alkaline phosphatase were obtained from Life Technologies, and terminal transferase and TdTase buffer were from Promega. All other chemicals were from Sigma. The SSPE buffer contained 10 mmol/L phosphate (pH 7.4), 0.18 mol/L NaCl, 1 mmol/L EDTA.

DNA extraction, PCR-single-strand conformation polymorphism, and direct manual sequencing

Tissue sections (10-μm thick) were cut from each frozen block. As discussed above, tumor tissue was microdissected according to histological evaluation. DNA was extracted using a nonorganic method (Oncor), and PCR was carried out using standard conditions. PCR products were diluted in denaturing loading dye, heated at 95 °C for 5 min, and flash-cooled on ice. Samples (4 μL) were loaded onto a 0.5 × MDE gel (FMC BioProducts) or a 10% glycerol gel, respectively; electrophoresis was at 5W for 16–20 h at room temperature. After electrophoresis, each gel was dried on a vacuum gel dryer and exposed to autoradiography film for 12–20 h.

Variant and wild-type bands for single-strand conformation polymorphism were cut out from the gels after alignment with the autoradiograph, and the DNA was eluted in 100 μL of doubly distilled H_2O at room temperature for 24 h and amplified by PCR. The PCR products were sequenced using the standard dideoxy chain termination approach, as recommended by the manufacturer (United States Biochemical). Samples were electrophoresed on an 8% sequencing gel at 75 W for 2–3 h. The gel was dried and exposed overnight at room temperature.

Sample preparation for chip analysis

Purified DNA (100 ng) was subjected to a multiplex-PCR where exons 2–11 were amplified simultaneously, using the reagents supplied by the manufacturer. Apart from the DNA, each PCR reaction contained 10 U of AmpliTaq Gold, PCR buffer II, 2.5 mmol/L $MgCl_2$, 5 μL of the primer set, and 0.2 mmol/L each dNTP. The reaction was carried out in a final volume of 100 μL. The PCR profile consisted of an initial heating at 95 °C for 10 min, followed by 35 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s, with a final extension step at 72 °C for 10 min. Forty-five μL of the PCR product was then fragmented by the addition of 0.25 units of fragmentation reagent [DNase I in 10 mmol/L Tris-HCl (pH 7.5), 10 mmol/L $CaCl_2$, 10 mmol/L $MgCl_2$, 500 mL/L glycerol] along with 2.5 U of calf intestine alkaline phosphatase, 0.4 mmol/L EDTA, and 0.5 mol/L Tris-acetate, and incubation at 25 °C for 15 min, followed by heat inactivation at 95 °C for 10 min. For labeling, 50 μL of the fragmented DNA was incubated at 37 °C for 45 min with 10 μmol/L fluorescein-N6-dideoxy-ATP, 25 U of terminal transferase, and TdTase buffer in a total volume of 100 μL, followed by heat inactivation at 95 °C for 10 min. The sample was hybridized to the chip in a volume of 0.5 mL containing 6× SSPE buffer, 0.5 mL/L Triton X-100, 1 mg of acetylated bovine serum albumin, 2 nmol/L control oligonucleotide, and the labeled DNA sample. Hybridization was done in an oven with constant agitation at 45 °C for 30 min. The chip was then washed on the wash station four times with 3× SSPE containing 0.05 mL/L Triton X-100. After washing, the chip was read by a confocal laser scanner, and the data were aligned and analyzed. A reference from the control DNA supplied was also analyzed. The reference was from the same PCR round and was measured on the same batch of chips.

Statistics

All data analysis was carried out using Corel Quattro spreadsheets, including calculation of means and standard deviations.

Results

Mutiplex-PCR reactions

The multiplex-PCR assay was highly reproducible under the described conditions. Fragments larger than the expected 10 exons were consistently present but did not interfere with the assay. Fragments could span several exons, including the corresponding introns, because some of the introns in the *TP53* gene are very small.

CHIP DESIGN

The Affymetrix p53 chip is manufactured photolithographically and contains an array with 65 536 squares with sides that are 50 μm each (Fig. 1). To each square are attached oligonucleotide probes that are ~20 bases long. Although several of the probes are control probes, used for alignment and quality control, an array of this size can analyze ~5 kb, thus being large enough to contain probes that can test the entire 1.2-kb coding sequence of the *TP53* gene, including a part of the intron/exon boundaries *(7)*. The analyzed nucleotides are each given a number, called the genechip position. The genechip positions are testing for the wild-type nucleotide, for substitution with the other three bases, and for a single base pair deletion (Fig. 1). Testing is done in both the sense and the antisense directions. This process is designated standard tiling. In addition to the control probes and the standard tiling, a third type of probe, called the alternate tiling, is included and tests known mutations and polymorphisms.

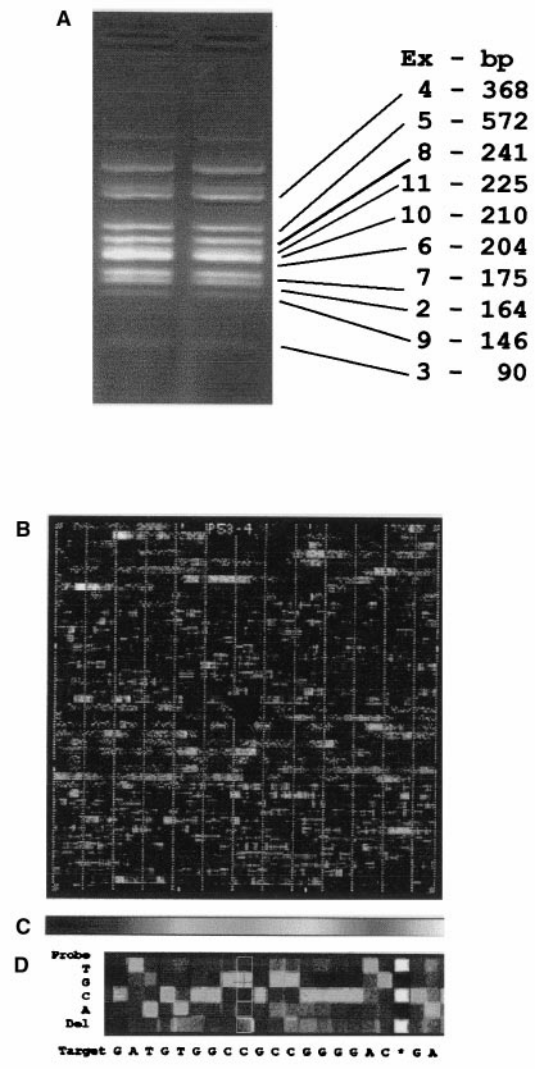

Fig. 1. Hybridization of PCR products to the Affymetrix GeneChip.

(*A*), example of the multiplex-PCR used to amplify all 10 exons detected by the chip. (*B*), the p53 chip after hybridization with a labeled multiplex-PCR product. The *checkerboard pattern* and the high-intensity *dotted vertical lines* are the control probes for alignment. (*C*), intensity scale, going from *black* (absence of signal) to *white* (maximum signal). (*D*), an enlarged section of the chip. This is an example of the standard tiling where the sequence can be read visually. Each *column* corresponds to a base in the sequence, and the *five rows* correspond to each of the four bases and to a single deletion. The column marked with * is from a column with alignment controls.

ANALYSIS OF CHIP SIGNALS

The analysis is done with a mixture detection algorithm, where the intensities of the sample are compared to the intensities of the reference (Fig. 2).

To test the performance of the chip, we analyzed DNA from 140 tumor samples from patients with bladder cancer. The DNA had previously been analyzed by single-strand conformation polymorphism followed by dideoxy sequencing. We found it unsatisfactory to use the results from the report generated by the software because many samples were reported to have several mutations simultaneously. These often occurred at certain places in the sequence, indicating that some genechip positions were more prone to noise problems than others. Additionally, it was unclear when the software decided that a score represented a mutation. To address the issue of background noise, we took all scores >0 from the score graphs and transferred the scores to a spreadsheet with 140 columns, each representing one sample, and 282 rows, 1 for each of the genechip positions where a score was found (Fig. 3).

The average (A) and standard deviation (S) for the background noise were calculated for each genechip position. Finally, data were plotted graphically along with lines showing A, $A + S$, and $A + 2S$ (Fig. 4A). Mutations confirmed by manual sequencing were not included in the calculation of the background statistics.

CHIP PERFORMANCE ON 140 BLADDER TUMOR SAMPLES

The manufacturer's description includes a 30-min hybridization step on the washing station. We did not find this approach satisfactory because it produced chips that appeared unevenly stained. The intensity was clearly higher at the part of the chip that was situated lowest in the wash station nearest to where the hybridization mix-

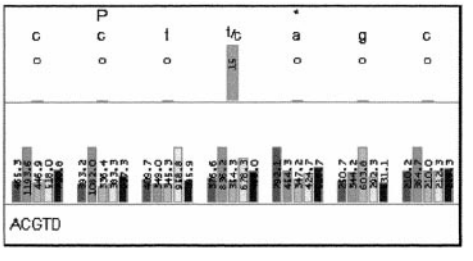

Fig. 2. Intensity measurements in each genechip position from the standard tiling represented as five columns representing the four bases (*A*, *C*, *G*, and *T*) and a single base deletion (*D*).

Above the standard tiling the row of single bars is the result of the analysis showing a score from the comparison with the reference intensities. At the top the sequence is listed.

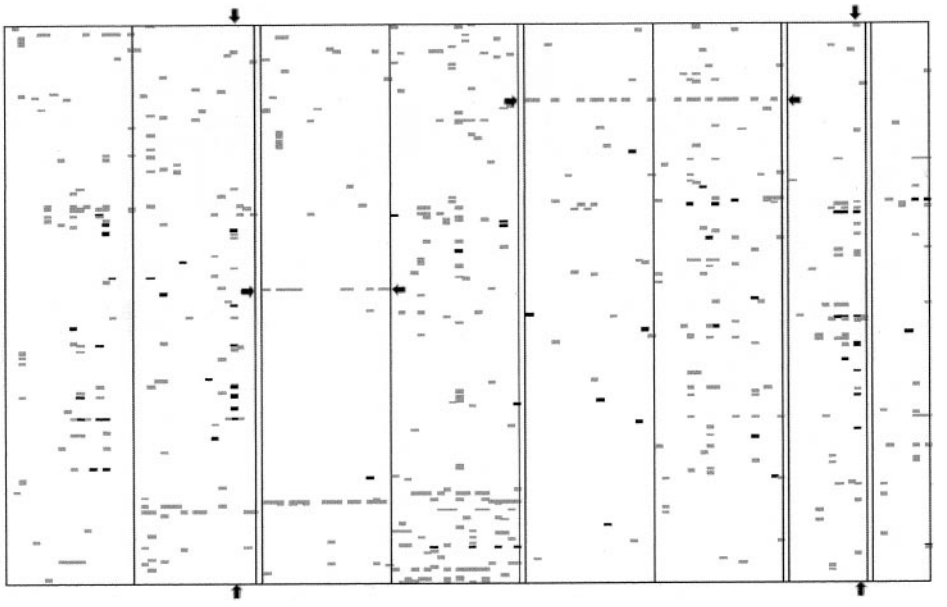

Fig. 3. The entire data set of the background scores.

Each *row* represents 1 of 251 genechip positions; each *column* represents 1 of 140 tumor samples. *Single vertical lines* divide groups of PCR rounds, and *double lines* divide different batches of chips. Scores from 1 to 9 are marked with *gray*, and scores >10 are in *black*.

ture comes into the chip. The agitation on the wash station is performed by pumping the mixture in and out from the bottom of the cartridge, and it was observed that the lower part of the chip was in contact with the mixture for a longer time than the upper part. Therefore, we chose to hybridize in an oven with constant agitation, which gave even staining (data not shown).

Data from 140 genechip experiments showed a large variation with respect to deviation from the references. Three of the chips had no scores and were therefore identical to the references, whereas the rest had scores between 1 and 26. Taking scores from the data set representing mutations confirmed by traditional sequencing enabled evaluation of the background noise. The size of the individual background scores from the whole data set showed an inverse exponential curve with no significant clustering, indicating a random distribution of the size of the noise, except for two very high scores (Fig. 4C).

In Fig. 4A, a graph representing the data from one genechip position shows the benefits of using the background statistics. The signal, with a score of 12, obtained from sample 16 is correctly ruled out by the $A + 2S$ value

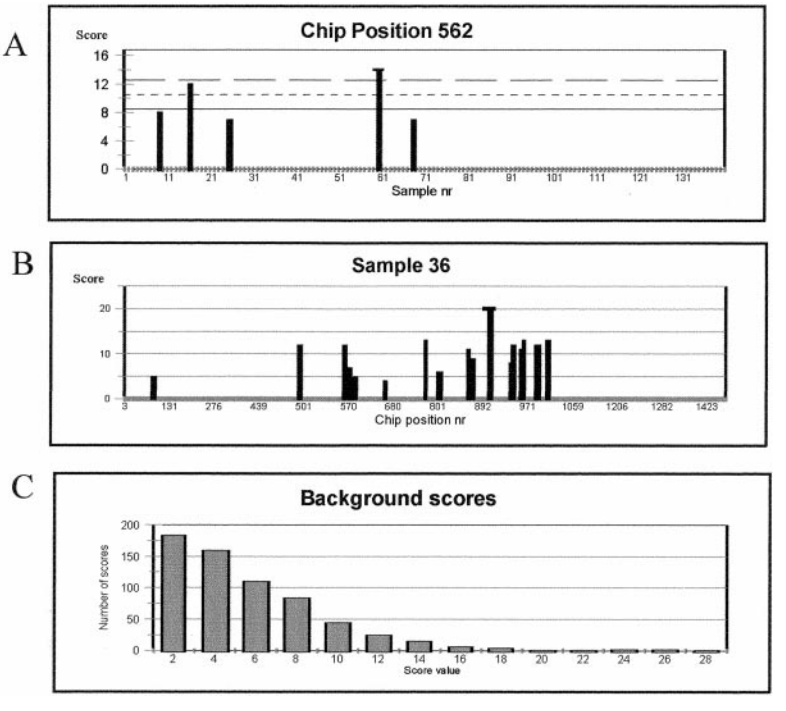

Fig. 4. Graphic plot of the background scores.

(*A*), a single genechip position from 140 chips with *A*, $A + S$, and $A + 2S$ indicated by *horizontal lines*. (*B*), 282 genechip positions from one chip. In both *A* and *B*, a confirmed mutation is marked with a *small horizontal bar*. (*C*), distribution of the size of the background scores.

of 12.6, whereas the confirmed mutation from sample 60, with a score of 14, is above that value. Of the 282 genechip positions, 31 had no background scores because only mutations were found at these genechip positions; whereas 130 had only one background score and could therefore be assigned only this value. Of the 251 positions with background scores, 203 had an $A + 2S$ value <10, whereas 48 had an $A + 2S$ value >10. The 31 genechip positions that contained mutations but had no background scores were included as true positives (Table 1, column Ch+/S+).

Three different methods for mutation calling were evaluated (Table 1). When we used the GeneChip report or a fixed cutoff, the chip method gave a large number of false mutation callings, whereas using the calculated cutoff values based on the background information from the whole data set gave a much better specificity.

Of the 11 mutations found by traditional sequencing but not by the chip, 4 were detected by the chip but with scores lower than the calculated $A + 2S$, and the other 7 were not detected at all by the chip. The latter did not seem to have any features in common.

The statistical data are attached to this article at http://www.clinchem.org/content/vol46/issue10, and they are also available at http://www.mdl.dk. Researchers are invited to submit data from their own p53 chip experiments to fpw@kba.sks.au.dk. The data will then be incorporated into the web page.

TITRATION EXPERIMENTS

We tested the sensitivity of the chip by analyzing a mixture of PCR products from two different cell lines with known mutations. The mutations were in different exons, so that the hybridization of the mutated oligomers would be on different probes and hence independent. In addition, we analyzed five of the mutations that had high scores in a mixture where they were present at a concentration of 20% each. The purpose of this analysis was to get an indication of to what extent the findings from the titration experiment could be generalized to other genechip positions. The titration experiments produced an S-shaped curve for both mutations (Fig. 5A). The scores were relatively stable in the range 20–80% but had a sharp rise or fall in the first and last 10%, respectively. Even at a content as low as 1%, the scores of the mutations were 10, at a genechip position where no background scores were found. The five mutations tested at 100% and 20% showed the same tendency, as seen in Fig. 5B.

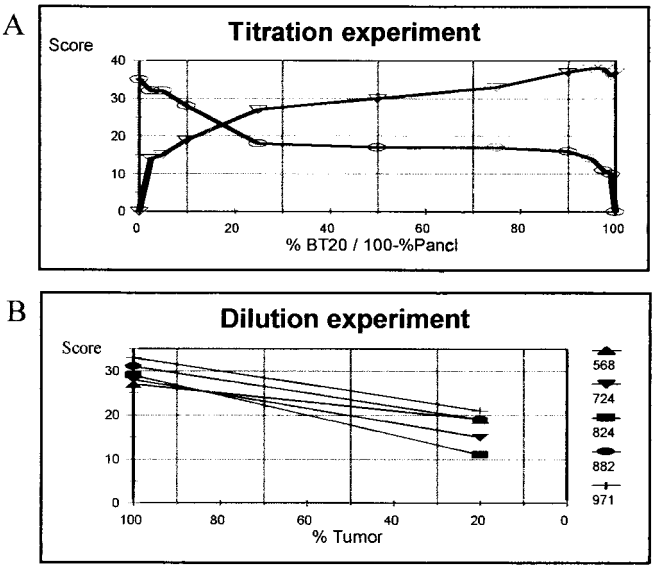

Fig. 5. Titration (*A*) and dilution (*B*) experiments on the Affymetrix GeneChip.

(*A*), plot of score value (*vertical axis*) vs percentage of mutated DNA (*horizontal axis*). ▽, strain BT20 in concentrations of 0–100%; ○, strain PancI in concentrations of 100–0%. (*B*), *symbols* denote the genechip positions where the mutations are found.

Although the slopes of the curves are different, it is obvious that none of them intersects zero.

Discussion

On an expression chip, there is some freedom of choice to find an oligonucleotide that hybridizes efficiently at a given temperature, although there are constraints because of specificity. On a sequencing chip, this freedom is absent because each position in the sequence must be tested. Although adjustments are possible, varying the length of the probe, large differences in both specificity and sensitivity must be expected. As our experiments have indicated, each probe should be regarded as a separate entity with its own binding properties and its own standardization requirements. Thus, information from one chip is not enough because cutoff values should be made for each genechip position, based on many experiments. Failure to do this has the consequence that many false mutations are reported. Ideally, one should have samples with mutations covering each base in the whole *TP53* gene, but this is not feasible because it would require >5000 different

Table 1. Performance of different mutation-calling methods compared with traditional sequencing.[a]

Method	Ch+/S+[b]	Ch+/S−	Ch−/S+	Ch−/S−	Sensitivity	Specificity
GeneChip Report	65	126	6	31	0.92	0.34
Fixed cutoff = 10	55	77	16	45	0.77	0.41
Calculated cutoff	60	10	11	65	0.84	0.86

[a] Insertions and deletions are not included in the table.
[b] Ch, chip sequencing; S, traditional sequencing.

mutations. On the other hand, our results do show that most of the genechip positions, namely all of the positions that did not have background scores, are positions of low noise. This does indicate that even a low score in one of these positions should be considered a positive mutation and preferably confirmed with another method until a cutoff value can be established. As mentioned earlier, we found 31 positions that had mutation scores but no background scores, with a score of 4 as the lowest encountered in this study. Other positions were error prone, and this was to some extent batch specific. Some positions seemed to be unable to detect mutations, and these poorly performing probes should be replaced by other probes. New chip chemistry based on synthetic nucleotides, or a mixture of conventional nucleotides and synthetic ones, might improve the performance of a given probe at a given stringency.

Ahrendt et al. *(5)* chose in their work to use a fixed cutoff value of 13, although some genechip positions have a possible maximum score of 7. Using the same approach in our work would have excluded one-half of the false positives, but it would also have created seven more false negatives, giving a sensitivity of 0.67 and a specificity of 0.59. This further supports the notion that probe-specific cutoffs are necessary.

The titration experiment showed an unexpectedly high sensitivity and indicates that the probes on the chip reaches saturation at a few percent of mutant DNA compared with wild-type DNA. Bearing in mind that there are two different probes involved for each mutation, one for the wild-type sequence and one for the mutated sequence, an obvious explanation for this S-shape can be offered. Each probe is contained in a square on the chip that carries between 10^6 and 10^8 oligonucleotides. Compared with the amount present in the sample used for hybridization—a few pmoles, or in the order of 10^{11}–10^{12} molecules—this is still a very low number. A detectable hybridization therefore occurs at a low concentration of sample, and because the algorithm presumably uses the intensities of both the affected squares, the increase as the mutation content approaches 100% is attributable to the fall in intensity of the wild-type sequence. It is a well-known problem when sequencing tissue samples that the tumor DNA is mixed with normal DNA from the surrounding tissue, to a lower or higher extent. Traditional sequencing methods usually have a threshold of ~30% because lower concentrations are obscured by the background noise *(8)*. Therefore, traditional sequencing will underestimate the number of missense mutations, whereas the chip method, for the above reasons, will tend to overestimate the number of missense mutations.

The chips did not detect any of the five frameshift mutations present in this study. Four of the mutations were insertions or large deletions, types of mutations that the chip is not designed to detect, but one single-base deletion should have been detected. The reason for this is not clear, but when one looks at the individual squares on the chip, it is a general feature that the squares detecting deletions often have some background intensity. If these positions generally are noise-prone, the software may be rejecting them as mutation callings.

Hacia *(3)* discussed the influence of a mutation on the signals from the neighboring cells. If 25-mer oligomers are used as probes, then a single base change mutation should not only have an influence on the hybridization of the probe squares testing for this particular position, but also on the nearest 24 positions. However, our titration experiment indicates that this probably is not a problem except in the rare case in which a homozygous mutation is present at a content of 100%. In our experiments, we did get positive scores at positions adjacent to mutations, scores that were not confirmed as mutations by traditional sequencing, but only in the homozygous cases. Avoiding this problem should be very easy because it would be enough to add a small fraction, e.g., 5%, of wild-type DNA to the sample.

The use of microarrays for sequencing is a promising new technology that in the future may offer sequencing as a diagnostic tool as rapid and cheap as other standard tests in the clinical biochemical laboratory. At this time, the total price per chip, including all necessary reagents, is approximately $190 per analysis, compared with approximately $120 for manual sequencing of the same 10 exons. The workload for manual sequencing, however, is much higher, thus making the chip an economically attractive alternative. Although much improvement is needed in the specificity and sensitivity, primarily concerning deletions and insertions, the technique as it is could be of great value as a first screening in certain types of cancer. Our results have shown that combining the data from several tests gives a tool that overcomes much of the imprecision in a single isolated chip. Furthermore, the detection limit is unprecedented. Having a sequencing method that detects a much lower percentage of mutant DNA could be of utmost importance because many tumors are heterogeneous, and a clinically important mutation could be detected early, before the tumor clone harboring the mutation could overgrow other clones.

This study was supported by grants from the Danish Cancer Society, Karen Elise Jensen Foundation, Fraenkels Fund, The University of Aarhus, and Aarhus County Research Fund.

References

1. Levin AJ. P53, the cellular gatekeeper for growth and division. Cell 1997;88:323–31.
2. Hollstein M, Sidransky D, Vogelstein B, Harrist CC. P53 mutations in human cancer. Science 1991;253:49–53.
3. Hacia JG. Resequencing and mutational analysis using oligonucleotide microarrays. Nat Genet 1999;21(1 Suppl):42–7.
4. Cronin MT, Fucini RV, Kim SM, Masino RS, Wespi RM, Miyada CG. Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. Hum Mutat 1996;7:244–55.

5. Ahrendt SA, Halachmi S, Chow JT, Wu L, Halachmi N, Yang SC, et al. Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. Proc Natl Acad Sci U S A 1999;96: 7382–7.
6. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays [Review]. Nat Genet 1999;21(1 Suppl):20–4.
7. Gunderson KL, Huang XC, Morris MS, Lipshutz RJ, Lockhart DJ, Chee MS. Mutation detection by ligation to complete n-mer DNA arrays. Genome Res 1998;8:1142–53.
8. Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, Chen SM. New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res 1997;15:25: 4500–4.