High-Throughput Genomic and Proteomic Analysis Using Microarray Technology

JOE XI HUANG, DOROTHY MEHRENS, RICK WIESE, SANDY LEE, SUN W. TAM, STEVE DANIEL, JAMES GILMORE, MICHAEL SHI, and DEVAL LASHKARI* 

Background: High-density microarrays are ideally suited for analyzing thousands of genes against a small number of samples. The next step in the discovery process is to take the resulting genes of interest and rapidly screen them against thousands of patient samples, tissues, or cell lines to further investigate their involvement in disease risk or the response to medication.

Methods: We used a microarray technology platform for both single-nucleotide polymorphisms (SNPs) and protein expression. Each microarray contains up to 250 elements that can be customized for each application. Slides contained either a 16- or 96-microarray format (4000 – 24 000 elements per slide), allowing the corresponding number of samples to be rapidly processed in parallel.

Results: Results for SNP genotyping and protein profiling agreed with results of restriction fragment length polymorphism (RFLP) analysis or ELISA, respectively. Genotyping analyses, using the microarray technology, on large sample sets over multiple polymorphisms in the NAT2 gene were in full agreement with traditional methodologies, such as sequencing and RFLP analysis. The multiplexed protein microarray had correlation coefficients of 0.82–0.99 (depending on analyte) compared with ELISAs.

Conclusions: The integrated microarray technology platform is adaptable and versatile, while offering the high-throughput capabilities needed for drug development and discovery applications.

© 2001 American Association for Clinical Chemistry

The ability to rapidly screen thousands of biological samples with a cost-effective and reproducible methodology is an extremely attractive prospect for drug developers. From compound selection through clinical trials, high-throughput technologies can accelerate the drug discovery and development process by bringing safer, more effective therapies to market. However, high-throughput technologies are usually single-point assays, capable of measuring only one analyte at a time. Microarrays have allowed researchers to assess multiple analytes in each assay, but are time-consuming to perform and cost-prohibitive to use in high-throughput screening. Currently, microarrays are used for two types of analyses: detection of mutations in specific genes and measurement of gene expression in diseased vs healthy cells.

In this report, we describe a microarray technology that addresses the need for high-throughput screening at an economic cost and is integrated in a manner such that the same technology can be used for genotyping, detection of differential gene expression, and quantitative protein expression. Each microarray contains 256 elements that can be customized to detect gene expression transcripts or single-nucleotide polymorphisms (SNPs). To expand the scope of this technology to examine proteins, a high-throughput protein microarray platform has also been developed, which simultaneously quantifies numerous protein analytes while using a minimal amount of sample. This is accomplished by the use of 16- or 96-well arrays printed on a single glass slide, with one sample hybridized per array. This technology permits rapid parallel screening of multiple proteins over thousands of samples, expediting the identification of novel drug targets and implementation of new diagnostic tests.

To meet the challenges of the new genomics era, we have developed a high-throughput microarray technology for genotyping, gene expression, and quantitative protein profiling. From array fabrication to data imaging and analysis, each step in the process is automated to

1 Nonstandard abbreviations: SNP, single-nucleotide polymorphism; NAT2, N-acetyltransferase 2; CCD, charge-coupled device; PBS, phosphate-buffered saline; IL, interleukin; TNF, tumor necrosis factor; and IFN, interferon.
permit high-throughput with consistent results. The studies described show the adaptability and versatility of the microarray platform, using data from NAT2 genotyping analyses and comparative studies between standard ELISAs and multiplexed protein microarrays.

**Materials and Methods**

**Array Fabrication**

**Genotyping arrays.** The 96-well Teflon®-coated glass slides were cleaned ultrasonically with a 1:10 dilution of detergent (Aquasonic cleaning solution; VWR) in warm distilled water followed by multiple rinses in distilled water and finally in absolute methanol. The slides were dried in an oven at 45 °C. Slides were then silanized by vapor deposition using 250 g/L 3'-glycidoxypropyltrimethoxysilane in dry xylene containing a catalytic amount of diisopropylethylamine (1). Silanization was performed at 85 °C under reduced pressure (25 inches of Hg) for 16 h.

Amine-labeled oligonucleotide capture probes were designed for seven SNP sites on the N-acetytransferase 2 (NAT2) gene: G191A, C282T, T341C, C481T, G590A, A803G, and G857A (Accession No. D10870). These were printed on silanized glass slides as described previously (2) by a custom microarray printer consisting of a 256-capillary print head mounted on a high-resolution X-Y-Z positioning robot. This print head can simultaneously deliver in a continuous manner up to 256 different probe solutions onto the glass substrate. Each capillary (100-μm diameter) originates from a 384-well plate contained in an oven at 45 °C. Slides were then silanized by vapor deposition with silane in xylene, using a modification of the method described in 1999 by Falipou et al. (3).

Two different concentrations of each analyte capture antibody in 0.1 mol/L carbonate buffer, pH 9.5, were printed on these silanized slides by a robotic arrayer. Table 1 shows the print pattern of the antibodies spotted and reagent control proteins.

**SNP Genotyping Assay**

The PCR was used to amplify each SNP on the NAT2 gene. Amplification was done in two steps, primary and secondary PCRs, to differentiate NAT2 from its family member gene NAT1. In the primary PCR, a pair of primers was designed to amplify the NAT2 gene encompassing all seven SNPs without amplifying NAT1. Amplifications were done in a 50-μL volume containing 400 nM primer, genomic DNA from blood (for other projects, 20 ng of sample DNA in solution is used), 200 μM each dNTP, 2.0 U of AmpliTaq Gold® DNA polymerase, and 1.5 mM MgCl₂. PCR was performed in a Perkin-Elmer 9600 Thermal Cycler, with the following program: 95 °C for 10 min; 35 cycles of 95 °C for 30 s, 55 °C for 1 min, 72 °C for 1.5 min; 72 °C for 7 min; and a final 4 °C hold. In the secondary PCR, primer pairs were designed to flank each of the seven polymorphic regions. Each primer was biotinylated at the 5’ end. PCR was conducted under conditions similar to those for the primary PCR except that 1 μL of a 1:1000 dilution of primary PCR product was used as template.

A 10-μL aliquot of the resulting multiplex PCR reaction was mixed with 50 μL of hybridization buffer containing 150 mmol/L sodium citrate and 5× Denhardt’s solution and then denatured at 95 °C for 10 min. A 50-μL aliquot of

| Table 1. Array map of individual capture antibodies.  
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNγ</td>
</tr>
<tr>
<td>800</td>
</tr>
<tr>
<td>IL-4</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>IL-6</td>
</tr>
<tr>
<td>400</td>
</tr>
<tr>
<td>IL-13</td>
</tr>
<tr>
<td>400</td>
</tr>
<tr>
<td>Fc-biotin</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

a Each antibody is spotted in triplicate. The print concentration (in mg/L) for each antibody is shown below the name. Fc-biotin and anti-rabbit IgG are printed as positive and negative controls, respectively, with Fc-Cy3 printed as fiduciary marker.
the denatured target solution was then applied directly on the microarray. The reaction was incubated at room temperature for 60 min. Two 15-min stringency washes with 35 mmol/L sodium citrate were used to remove single-base mismatches, leaving labeled amplimers forming perfect-match duplexes with allele-specific probes.

A 1:1000 dilution of streptavidin–alkaline phosphatase conjugate (Roche) was applied to the microarrays for 30 min to bind with biotin-labeled targets. After the excess was washed away, an enzyme-linked fluorescence detection solution (Molecular Probes) was applied to the microarrays and incubated for an additional 30 min. Arrays were imaged after removal of excess detection solution. Fig. 1A is a schematic description of the genotyping microarray assay.

MEASUREMENT OF CYTOKINES USING ELISAS AND PROTEIN MICROARRAYS

ELISA reagent sets were purchased from BD PharMingen and BioSource, and analyses were performed according to the manufacturers’ instructions and specifications. Arrays were rinsed three times with phosphate-buffered saline (PBS) containing 0.5 mL/L Tween 20, followed by blocking with casein in PBS (Pierce) for 1 h at room temperature. Arrays were rinsed and incubated for 2 h at 37°C after application of 20 μL of antigen solution in PBS containing 100 mL/L fetal bovine serum. Biotinylated detector antibodies were applied to arrays at appropriate concentrations and incubated for 1 h at 37°C. The arrays were rinsed, and streptavidin-R-phycoerythrin in PBS containing 100 mL/L fetal bovine serum was added. Slides were incubated in the dark for 1 h, followed by washes with PBS containing 0.5 mL/L Tween 20. Slides were scanned after a final wash with PBS.

IMAGING AND ANALYTICAL EVALUATION OF DATA

SNP genotyping. Microarrays were imaged with a custom CCD-based imager (2). Custom software was used to quantify the resulting fluorescent images and to score the data for the presence or absence of each allele in each of the hybridized samples. Fig. 1B is representative of hybridization signals from three genotypes of a SNP.

PROTEIN MICROARRAYS

After drying, slides were scanned in a ScanArray® scanner (Packard Instruments) with a 532 nm laser. The location of each analyte element was outlined using OptiQuant gridding software. The same software was also used to measure the signal from each element. The background, calculated as the median of pixel intensities from the local area around each element, was subtracted from the average pixel intensity within each element to determine signal for each analyte element.

Results and Discussion

PHARMACOGENOMIC APPLICATIONS USING MICROARRAYS FOR SNP ANALYSIS

Genetic variations present in most drug-metabolizing enzymes alter response to drug therapy (4). We studied a phase II drug-metabolizing enzyme, NAT2. The gene that codes for this enzyme is highly polymorphic, with seven SNPs (G191A, C282T, T341C, C481T, G590A, A803G, and G857A) identified in the coding region that alter the enzyme’s ability to acetylate certain therapeutics (5).

We used the NAT2 array to genotype blood samples (obtained from The Houston Blood Center, Houston, TX) from a Population Diversity Panel (94 each of the African-American, Asian-American, and Caucasian populations and 91 of the Hispanic population; all self-identified). Genotype frequencies for the seven NAT2 SNPs were calculated from the array-based genotype calls (Table 2). Consistent with previous reports, the prevalence of each polymorphism varied greatly among populations (6–8). The variant form NAT2 G191A was predominant in the African-American population tested, but had a low frequency in the Hispanic population. This allele was absent in the Asian and Caucasian subgroups tested in this study. The G857A allele, although relatively uncommon in the general population, was present at the highest frequency in the Asian population.
frequencies in the Asian subgroup. Variant alleles T341C, C481T, and A803G were less frequent in the Asian population than in other ethnic groups, whereas the variant G590A was more prevalent in this group. Other trends in this study showed that the Caucasian population tested had higher frequencies of the T341C and C481T alleles, both consistent with the *5A and *5B slow acetylator phenotypes, compared with the other ethnic subgroups. This finding is consistent with past studies that showed that the Caucasian population, in general, has a higher number of slow acetylators than other ethnic groups (6).

Although additional haplotype information is needed in some cases to accurately predict acetylator phenotype based on NAT2 genotype, these results indicate that the high-throughput microarray platform may be a useful tool for rapidly determining the frequency of genotypes over a large population of samples. With more than 1 million biallelic SNPs proposed as markers for disease gene identification (9, 10), advanced genotyping technologies will be a driving force in large-scale pharmacogenetic studies, as needed in clinical trials or diagnostics. This microarray-based platform has advantages over traditional genotyping methodologies, allowing parallel analyses of large numbers of samples over multiple polymorphisms. Additional investigations using high-throughput microarrays to genotype polymorphisms for drug-metabolizing enzymes in diverse populations for drug response and disease association studies are currently underway.

**PROTEIN MICROARRAY APPLICATIONS**

Initial experiments addressed antibody specificity and potential cross-reactivity between cytokines in a multiplex assay. We used eight unique detector antibodies either alone or as a pool after incubation of the array with individual antigens, at concentrations of 10–1000 ng/L. Under all experimental conditions tested, similar antibody specificities were obtained between individual and multiplexed detector antibodies with no cross-reactivity seen in the multiplexed assay (data not shown). To assess the accuracy and precision of quantification and the detection limits of the platform, the multiplexed protein microarrays were compared with standard ELISAs.

Cytokine calibration curves obtained using conventional ELISA methods and the multiplex protein microarray were compared. Calibration curves for the microarray assay were generated using both a single cytokine with its appropriate detector antibody and multiple cytokines with pooled detector antibodies. Fig. 2 shows a representative calibration for one of the cytokines on the array (other cytokines showed similar results; data not shown) and suggests a sensitivity similar to or greater than that of ELISAs.

The ability of the multiplexed protein microarray to simultaneously quantify cytokines in complex cell-derived supernatants was determined. Jurkat cells were stimulated with a combination of 2 mg/L phytohemagglutinin and 10 μg/L phorbol-12-myristate-13-acetate,
and the culture supernatants were harvested at various time points after stimulation. The supernatants were analyzed by standard ELISA as well as by multiplexed protein microarrays. Fig. 3 shows comparative data for interleukin-2 (IL-2), tumor necrosis factor-α (TNF-α), and interferon-γ (IFN-γ). Other cytokines were undetectable. Control applications of dimethyl sulfoxide also had no detectable cytokine (data not shown).

In conclusion, this study identifies and discusses two applications of a high-throughput, microarray technology platform. Other applications include compound library screening, lead optimization and prioritization, toxicity profiling, diagnostics, and other areas of clinical profiling. The high-throughput capability of this technology platform offers the ability to avoid bottlenecks in key steps of the drug discovery and development process, as well as low-cost alternatives to current genomic and proteomic methodologies.

We would like to acknowledge Rick Eggers and Bonnie Iverson for expert contributions in the design of the NAT2 genotyping assay; Brandon Cook, Audrea Haynes, and Orlando Logan for technical assistance in the NAT2 genotyping analysis; and Krishnanand Kumble and Mary Jane Cunningham for expert discussion and guidance in protein analysis. We thank Bobbie Ireland for assistance in preparing the manuscript.

References