Anal. Chem. **2010,** *82,* 8551–8557

Label-Free Prehybridization DNA Microarray Imaging Using Photonic Crystals for Quantitative Spot Quality Analysis

S. George,[†,‡] I. D. Block,[‡,§] S. I. Jones,[∥] P. C. Mathias,[†,‡] V. Chaudhery,[‡,§] P. Vuttipittayamongkol,[§] H.-Y. Wu,[‡,§] L. O. Vodkin,[∥] and B. T. Cunningham*[,†,‡,§]

Department of Bioengineering, University of Illinois at Urbana−Champaign, 1304 West Springfield Avenue, Urbana, Illinois 61801, Micro and Nanotechnology Laboratory, University of Illinois at Urbana−Champaign, 208 North Wright Street, Urbana, Illinois 61801, Department of Electrical and Computer Engineering, University of Illinois at Urbana−Champaign, 1406 West Green Street, Urbana, Illinois 61801, and Department of Crop Sciences, University of Illinois at Urbana−Champaign, 1102 South Goodwin Avenue, Urbana, Illinois 61801

Technical variability during DNA capture probe printing remains an important obstacle to obtaining high quality data from microarray experiments. While methods that use fluorescent labels for visualizing printed arrays prior to hybridization have been presented, the ability to measure spot density using label-free techniques would provide valuable information on spot quality without altering standard microarray protocols. In this study, we present the use of a photonic crystal biosensor surface and a high resolution label-free imaging detection instrument to generate prehybridization images of spotted oligonucleotide microarrays. Spot intensity, size, level of saturation, and local background intensity were measured from these images. This information was used for the automated identification of missed spots (due to mechanical failure or sample depletion) as well as the assignment of a score that reflected the quality of each printed feature. Missed spots were identified with >95% sensitivity. Furthermore, filtering based on spot quality scores increased pairwise correlation of posthybridization spot intensity between replicate arrays, demonstrating that label-free spot quality scores captured the variability in the microarray data. This imaging modality can be applied for the quality control of printed cDNA, oligonucleotide, and protein microarrays.

DNA microarrays are an important tool for the highly parallel study of expression levels of multiple genes of interest. However, microarray data has been shown to lack reproducibility,[1] which in part is due to variability in array printing.[2] Custom arrays that are printed in academic laboratories offer greater content flexibility and lower cost than commercial arrays but are especially susceptible to variable print quality due to the absence of consistent quality analysis.[3] Without robust quality control measures that account for this source of noise, the use of fluorescence spot intensity information for inferring gene expression levels, the underlying principle of microarray analysis, can be inaccurate.

A high quality printed microarray slide possesses a uniform coating of a functional group that yields spots bound to the slide that are of desired density, shape, and size. Insufficient amounts of surface-bound probes can result in the underestimation or even complete failure of detecting gene expression.[2,4] In order to account for the variability that primarily stems from variable DNA probe deposition and retention on the slide surface, a few label-based methods have been devised for the visualization of printed slides prior to their experimental use. In one such approach, a labeled oligonucleotide target specific for a common vector sequence found in all immobilized probe spots has been used to visualize array spots.[4] While this method provides information on each spot's density and morphology, the analysis is performed on a single test slide sacrificed from each print batch which may not be equivalent to the slides used for the experiment. In another method, the probe spots are stained with an ssDNA-binding dye like SYBR Green II after printing.[5] After visualization and quality analysis, these slides are destained before hybridization with the labeled target. The effect of this destaining process on subsequent slide performance can be a concern. A third approach that uses probes with cyanine-compatible, fluorescein-labeled primers for visualizing cDNA-based microarrays and generating quantitative assessments of spot quality has been reported.[6] However, the application of this method to spotted oligonucleotide arrays requires the labeling of each probe sequence, which is cost prohibitive. As an alternative, the use of a fluorescein-labeled

* Corresponding author. Address: Micro and Nanotechnology Laboratory, University of Illinois at Urbana−Champaign, 208 North Wright Street, Urbana, Illinois 61801. E-mail: bcunning@illinois.edu. Phone: 217-265-6291.

† Department of Bioengineering.

‡ Micro and Nanotechnology Laboratory.

§ Department of Electrical and Computer Engineering.

∥ Department of Crop Sciences.

(1) Chuaqui, R. F.; Bonner, R. F.; Best, C. J. M.; Gillespie, J. W.; Flaig, M. J.; Hewitt, S. M.; Phillips, J. L.; Krizman, D. B.; Tangrea, M. A.; Ahram, M. *Nat. Genet.* **2002**, *32*, 509–514.

(2) Wang, X.; Jia, S.; Meyer, L.; Xiang, B.; Chen, L. Y.; Jiang, N.; Moreno, C.; Jacob, H. J.; Ghosh, S.; Hessner, M. J. *BMC Bioinf.* **2006**, *7*, 378.

(3) Larkin, J. E.; Frank, B. C.; Gavras, H.; Sultana, R.; Quackenbush, J. *Nat. Methods* **2005**, *2*, 337–344.

(4) Yue, H.; Eastman, P. S.; Wang, B. B.; Minor, J.; Doctolero, M. H.; Nuttall, R. L.; Stack, R.; Becker, J. W.; Montgomery, J. R.; Vainer, M. *Nucleic Acids Res.* **2001**, *29*, e41.

(5) Diehl, F.; Grahlmann, S.; Beier, M.; Hoheisel, J. D. *Nucleic Acids Res.* **2001**, *29*, e38.

(6) Hessner, M. J.; Wang, X.; Hulse, K.; Meyer, L.; Wu, Y.; Nye, S.; Guo, S. W.; Ghosh, S. *Nucleic Acids Res.* **2003**, *31*, e14.

10.1021/ac101551c
Published on Web 09/21/2010

"tracking" oligonucleotide, introduced in the printing buffer at a low concentration for indirectly monitoring spot fidelity, was developed.[7] While this method can identify mechanical misses or spots where the target-specific oligonucleotide was present at a very low concentration, it is not clear if the signal dynamic range from the tracking oligonucleotide is sufficient for capturing the variability in concentration of the target-specific oligonucleotide for the remaining spots.

Such label-based methods require the modification of immobilized probes or a change in typical microarray protocols, limiting their widespread adoption. Here, we report on the use of a photonic crystal (PC) biosensor-based label-free (LF) detection method for generating prehybridization images of mass density for the quantitative assessment of spot quality. LF sensing is achieved by monitoring changes in the optical resonance condition of the PC as biomolecules with dielectric permittivity greater than the surrounding medium (air) are adsorbed to the surface. Using a PC surface in place of a standard glass microscope slide to obtain spot information, this method can be applied to spotted cDNA, oligonucleotide, and protein microarrays alike, with no modification to standard spotting or hybridization protocols.

The PC is composed of a periodically modulated low refractive index nanoscale structure that is then coated with a high index dielectric layer, shown schematically in Supplementary Figure S1 (Supporting Information). Such a device behaves as a narrow band optical filter with nearly 100% reflection (or correspondingly, ~0% transmission) when illuminated at a particular wavelength and incident angle combination, known as the resonant condition. For all other combinations, nearly 100% of the light is transmitted through the device. Since the structure consists of a grating with a subwavelength period, only zeroth order reflected and transmitted modes are propagated while higher order diffraction modes are cut off. To excite the PC resonance while illuminating the device at a fixed wavelength, one modulates the angle at which the light is incident on the surface and generates a measurement of the PC transmission efficiency as a function of incidence angle. Only a single incidence angle fulfills the resonance condition and nearly 0% transmission is observed at this angle. The resonant angle shifts to higher angles as the refractive index near the surface of the PC is increased by the adsorption of biomolecules. In this manner, the shift in the resonant angle can be used to quantify the density of biomolecules deposited on the surface. Because resonant light is not allowed to propagate laterally across the PC surface, the angle of minimum transmission (AMT) at a particular location is independent of adjacent locations, so it is possible to generate high resolution spatial maps of adsorbed biomolecular density by measuring the AMT shift as a function of position on the PC surface. As shown in Supplementary Figure S2 (Supporting Information), the PC surfaces used in this study were designed and fabricated to provide a resonance at wavelength of $\lambda = 633$ nm for an incident angle of ~2 degrees. Using plastic-based nanoreplica molding techniques described previously,[8] a continuous uniform PC surface is prepared on a flexible polyester substrate in a roll-to-roll manufacturing process and applied with adhesive to a standard 1×3 in.^{2} glass microscope slide. The entire microscope slide surface is composed of the PC surface.

LF images were captured using a custom high resolution imaging microscope described previously.[9] This instrument uses an angle-tunable, single-wavelength ($\lambda = 632.8$ nm) laser source to illuminate (through the substrate) a PC that has been prepared with immobilized microarray capture spots. While the PC is illuminated from below, a microscope objective and CCD camera gather a sequence of images of transmitted intensity from above the PC, where a sequence of images are rapidly gathered over a range of incident angles using small increments. From this stack of images, the resonant angle can be determined on a pixel by pixel basis by fitting the transmission versus angle data to a Lorentzian curve such that a label-free image of surface bound molecular density can be constructed. This spatial map of resonant angles, called the "AMT image" is the LF image of the microarray. A schematic diagram of the LF microscope is shown in Supplementary Figure S3, Supporting Information.

We printed a custom 70-mer oligonucleotide array on PC surfaces and subsequently captured LF images. The arrays were then hybridized with a cyanine-5 labeled target, and fluorescence images were obtained. The LF images were used to generate quantitative information on spot characteristics such as morphology, surface DNA retention, and background. Using LF spot density information, a method was developed for the automated identification of missing spots in LF images in order to differentiate false negatives (due to a mechanical miss during printing or problem with the source well) from true negatives (due to the absence of gene expression). Next, to capture the variation in the probe density immobilized on the PC, a previously published microarray spot quality analysis method,[2] designed for fluorescence prehybridization images, was adapted to our LF images. This algorithm was used to generate a composite quality score for each spot based on the following metrics: signal to background ratio, spot size, and level of pixel saturation. These quality metrics have been shown to capture variability in the microarray data wherein gene expression measurements from spots with high quality scores yield less variation than those made from low scoring spots.[6,10] We evaluated the use of these composite scores for quality filtering by comparing the reproducibility of unfiltered fluorescence hybridization data with data filtered on the basis of spot scores. We found that filtering based on spot quality scores increased pairwise correlation of posthybridization spot intensity between replicate arrays, demonstrating that label-free spot quality scores captured the variability in the microarray data.

MATERIALS AND METHODS

Device Fabrication and Characterization. The PC devices used, shown schematically in Figure S1 (Supporting Information), were fabricated using a nanoreplica molding process as described previously.[8] Briefly, a silicon "master" wafer with a negative volume image of the one-dimensional periodic grating structure was made using deep-UV lithography and reactive ion etching. A liquid UV-curable polymer (UVCP) was dispensed on a polyethylene tetraphthalate (PET) sheet, and the pattern from the master

(7) Hessner, M. J.; Singh, V. K.; Wang, X.; Khan, S.; Tschannen, M. R.; Zahrt, T. C. *BMC Genomics* **2004**, *5*, 12.
(8) Cunningham, B.; Lin, B.; Qiu, J.; Li, P.; Pepper, J.; Hugh, B. *Sens. Actuators, B* **2002**, *85*, 219–226.
(9) Block, I. D.; Mathias, P. C.; Ganesh, N.; Jones, S. I.; Dorvel, B. R.; Chaudhery, V.; Vodkin, L. O.; Bashir, R.; Cunningham, B. T. *Opt. Express* **2009**, *17*, 13222–13235.
(10) Wang, X.; Ghosh, S.; Guo, S. W. *Nucleic Acids Res.* **2001**, *29*, e75.

was transferred to the polymer with a roller. After the polymer was cured using a high intensity UV lamp, the replica was peeled away from the master. Next, an index matched 300 nm thick layer of SiO_2 was deposited via sputtering to act as a spacer between the autofluorescent UVCP layer and a high refractive index layer of TiO_2. A 160 nm thick layer of TiO_2 was finally sputtered on the device. The completed devices were cut into 25 mm × 75 mm sections and attached to glass microscope slides using an optically transparent adhesive. Optical characterization of the PC slides was performed by illumination with polarized, collimated white light and collection of transmitted light into a UV−visible light spectrometer (Ocean Optics). To measure the dispersion of the devices and determine the angle of resonance, PC slides were mounted on an angle-adjustable stage and rotated under illumination with broadband light polarized with an electric field vector oriented perpendicular to the grating lines. The transmission spectrum of a device as a function of angle is shown in Figure S2, Supporting Information. The PC slides used in this study showed good spectral uniformity of the narrow excitation resonance over their area with a maximum observed resonant angle of $\sigma_{\text{within-PC}} = 0.12°$ (corresponding coefficient of variance = 21.1%). Furthermore, we obtained label free images of silanized slides prior to printing. Figure S4 (Supporting Information) shows line profiles taken along the long axis of three PC slides used in this study.

Surface Chemistry and Array Printing. Due to its documented low autofluorescence,[11] an epoxysilane-based surface chemistry was used to functionalize the PC slides. All slides were first cleaned in an O_2 plasma system and then incubated overnight with 3-glycidoxypropyltrimethoxysilane at 185 mTorr. Oligonucleotides were printed on the slides using a Genetix QArray^2 contact pin-spotter. A set of 192 70-base oligonucleotides consisting of soybean genes annotated as storage proteins, cell wall proteins, transcription factors, and other genes of interest were spotted on the slides. These 192 oligonucleotides are part of a larger set of 38 400 oligonucleotides detailed previously.[12] A total of six slides from two separate surface chemistry and printing batches were used in this study. The data pooled from all slides consisted of 32 640 spots of which 1300 were buffer blank spots used as negative controls.

Label-Free Image Acquisition and Processing. After rinsing any unbound DNA, label-free images of the printed PC slides were obtained using a custom built imaging microscope. A detailed description of the instrument and its operation has been presented previously[9] and is briefly discussed here. Label-free detection was performed by imaging the transmission of a transverse magnetic (TM) polarized (electric field perpendicular to the PC grating lines) and collimated 35 mW helium−neon $\lambda = 632.8$ nm laser through the PC as a function of laser incidence angle. Images were captured using a 16-bit EM-CCD (Hammamatsu, Japan), and the incidence angle was computer controlled and scanned for a fixed range of angles about the expected resonance location, typically between 0° and 3°, in increments of 0.01°. This transmission versus angle data was then fit per pixel to generate a label-free image at a pixel resolution of 8 μm. In this image, each pixel was assigned the fitted value for the angle of resonance. Using a 2× 0.06 NA objective (Olympus), the imaging system gathers an AMT image from a 16 mm^2 area. The system is capable of generating concatenated images of an entire microscope slide with a computer-controlled motion stage and software that can stitch together images from adjacent regions. Signal segmentation and spot intensity calculations of the label-free images were carried out using GenePix Pro 6.1. Spot detection settings were chosen with a large scope (33−300%) of feature size, and without a minimum feature threshold, which means that no spots were eliminated at this stage. This raw data for all spots was used for further analyses.

Sample Preparation and Hybridization. Soybean seeds (*Glycine max* cultivar Williams) with fresh weight between 100 and 200 mg were dissected to separate the cotyledon and then lyophilized. Using previously published protocols,[12] total RNA was extracted from the cotyledons. The RNA sample was purified using a Qiagen RNeasy kit and labeled using a direct-label procedure with Cy5-dUTP. Approximately 40 μg of total RNA was used per slide. Slides were blocked prior to hybridization with bovine serum albumin, hybridized at 42 °C overnight with the labeled sample, and then washed.

Fluorescence Image Acquisition and Processing. Posthybridization, all slides were scanned with a commercial confocal microarray scanner (LS Reloaded, Tecan) with a TM polarized laser ($\lambda = 632.8$ nm) at normal incidence. All slides were scanned at the same photomultiplier tube (PMT) gain and at a resolution of 10 μm. Signal segmentation and spot intensity calculations of fluorescence images were carried out using GenePix Pro 6.1. Spot detection settings were chosen with a large scope (33−300%) of feature size and without a minimum feature threshold. The raw data for all spots except those flagged as "not found" by GenePix were used for further analyses.

RESULTS AND DISCUSSION

A total of six PC slides from two separate surface chemistries and array print batches were used in this study. Prehybridization LF images of two subarrays from a representative slide are shown in Figure 1a,b, and line profiles taken through these images are shown in Figure 1c,d. It can be observed that areas where the probe DNA has been immobilized produce a measurable increase in the resonant angle.

Each array was designed to include buffer blank spots as negative controls in predetermined locations across the slide. However, during visual inspection of the LF images, we noticed the presence of additional unintended missed spots (caused due to a mechanical failure or insufficient material in the source well) on slides from one of the print batches. Figure 1a shows a subarray containing intentional blank spots while Figure 1b shows a different subarray on the same slide that contains missed spots. It is important to systematically identify and eliminate all such instances that result in false negatives in the posthybridization expression data that is analyzed. Thus, we sought to automate the identification of all missing spots in LF prehybridization images. First, all LF images were quantitated using the Genepix Pro 6.1 software (Axon Instruments) to obtain spot intensity, background intensity, and spot size information. Each spot was then assigned an intensity score, $q_{\text{intensity}}$ as follows:

(11) Taylor, S.; Smith, S.; Windle, B.; Guiseppi-Elie, A. *Nucleic Acids Res.* **2003**, *31*, e87.
(12) Gonzalez, D. O.; Vodkin, L. O. *BMC Genomics* **2007**, *8*, 468.

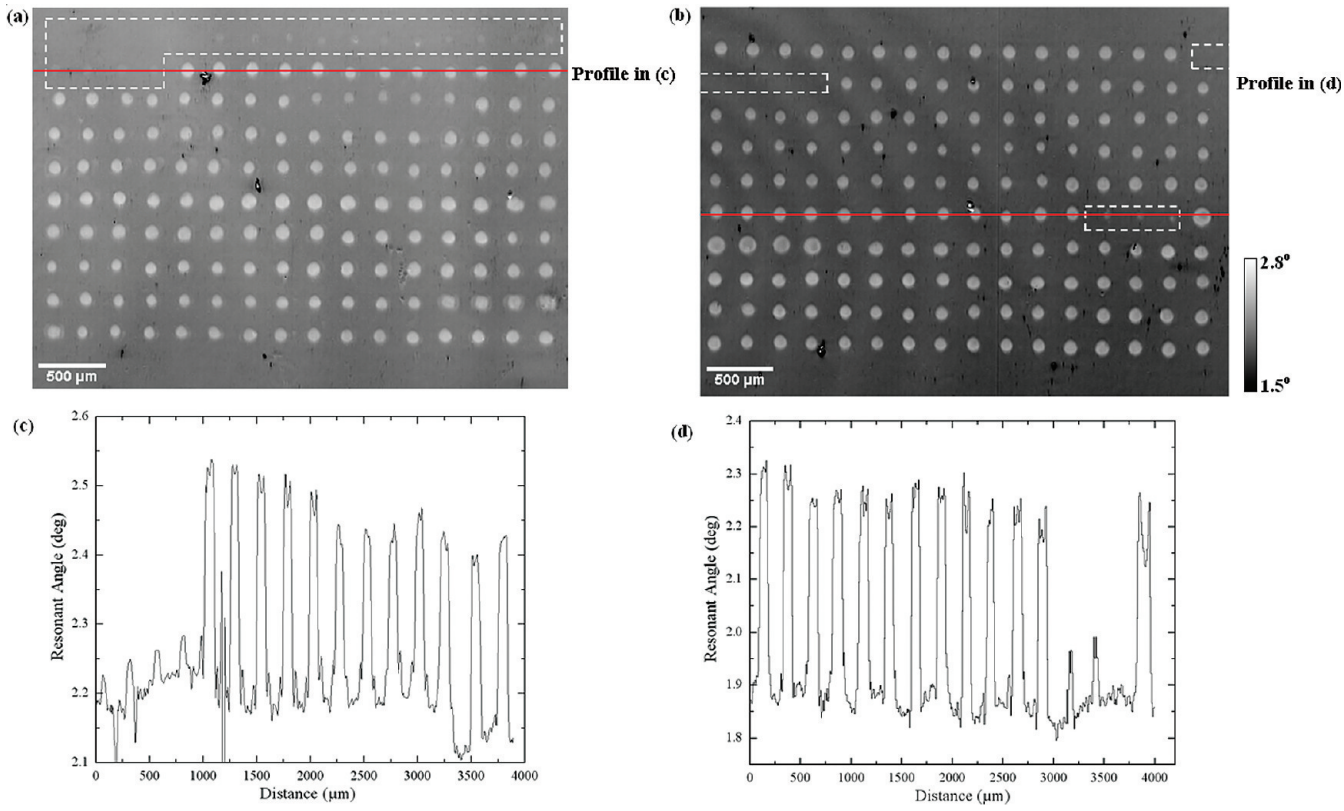


Figure 1. LF images of two subarrays from the same PC slide are shown in (a) and (b). The white dashed box in (a) denotes the location of a set of 20 intentional blank spots. A line profile running through a row containing three blank spots followed by 13 probe spots is shown in (c).The white dashed boxes in (b) denote locations of unintentional missed spots. A line profile running through a row in this subarray is shown in (d). The difference in background signal between subarrays indicates spatial variations in the resonance angle that is a function of device and surface chemistry nonuniformities. Dark specks in the images are due to device imperfections.

$$q_{\text{intensity}} = (q_{\text{sig-bkg}} \times q_{\text{sig-sig0}})^{0.5} \quad (1)$$

$$q_{\text{sig-bkg}} = \frac{S_{\text{mean}}}{S_{\text{mean}} + B_{\text{mean}}} \quad (2)$$

$$q_{\text{sig-sig0}} = \frac{S_{\text{mean}}}{S_{\text{mean}} + S_{\text{mean0}}} \quad (3)$$

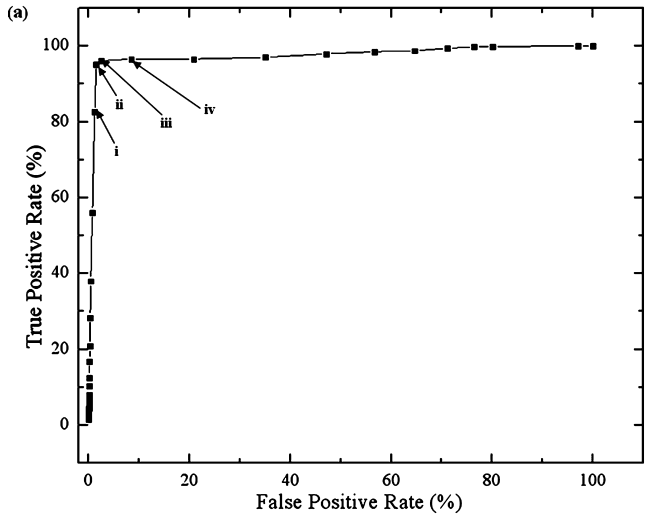
where S_{mean} = mean intensity of all pixels in spot, B_{mean} = mean intensity of all pixels in spot local background, and S_{mean0} = mean intensity of all spots in the corresponding subarray.

The local background region is defined as the annular area around a spot that does not include any neighboring spots and has an outer diameter 3 times that of the spot. All pixels in this region contribute to the local background signal.

For each spot, $q_{\text{sig-bkg}}$ is a measure of its density and $q_{\text{sig-sig0}}$ is a measure of its intensity relative to all spots in the subarray, and both measures are values between 0 and 1. Our initial attempt to use only $q_{\text{sig-bkg}}$ to screen for missing spots was aborted due to the observed high false positive rate. This was due to spatial variations in the device resonant angle due to measurable surface chemistry gradients that increased spot background intensity, consequently reducing $q_{\text{sig-bkg}}$. To circumvent this problem, we developed the second measure, $q_{\text{sig-sig0}}$, that compares the spot intensity to the mean intensity of all spots in the block. Using this measure, good spots located in regions of high local background were not penalized. A $q_{\text{intensity}}$ value approaching 1 denotes a predicted present spot while a value approaching 0 denotes a predicted absent spot.

Independently, images were visually inspected, and each spot was manually classified into two categories (present, absent) by two users. A spot that was classified as absent by either user was assigned to the "absent" category, and all remaining spots were assigned to the "present" category.

The $q_{\text{intensity}}$ for spots from all six slides were pooled, and the prediction performance based on $q_{\text{intensity}}$ was evaluated by receiver operating characteristic (ROC) analysis. The ROC curve shown in Figure 2a describes the ability of the parameter $q_{\text{intensity}}$ to discriminate between spots that are present and absent. Using this curve, the threshold $T_{\text{m}} = 0.50$ was identified to have the highest prediction power (that is maximum specificity and sensitivity). Spots with $q_{\text{intensity}}$ below T_{m} were predicted as absent whereas the remaining spots were predicted as present. At $T_{\text{m}} = 0.50$, the sensitivity (absent spots predicted as absent) was 95.2% and the specificity (present spots predicted as present) was 98.5%. Of all identified true positives, 91.7% were intentional blank spots and 8.3% were unintentional missed spots. At this threshold, 96.5% of all intentional blank spots were identified. A density plot of the $q_{\text{intensity}}$ values for spots manually classified as absent and present is shown Figure 2b. There exists a partial overlap in $q_{\text{intensity}}$ values between the two classes of spots. The threshold value T_{m} chosen represents the separation point between these categories that maxi-



Data point	$q_{intensity}$ threshold value	True Positive Rate	False Positive Rate
i	0.49	82.7	1.2
ii	0.5	95.2	1.5
iii	0.51	96.1	2.5
iv	0.52	96.6	8.5

Figure 2. (a) ROC curve demonstrates the ability to discrimination between spots that are absent or present on the slide using $q_{intensity}$. Four threshold values, selected for their high true positive rate (or sensitivity) and low false positive rate (or 1-specificity) as indicated in the figure are presented in the table. By selecting a $q_{intensity}$ threshold of 0.5 (where spots with $q_{intensity} < 0.5$ are predicted as absent), the predication sensitivity and specificity were 95.2% and 98.5%, respectively. As shown in (b), there is a partial overlap between the $q_{intensity}$ distributions for the two classes and the threshold is the separating point between the classes.

mizes overall prediction accuracy. Our array was designed to contain a total of 40 technical replicates per slide located on different subarrays in sets of five continuous replicates. We found that, in many cases of unintentional missed spots, all five replicate spots in a subarray were missed. For all instances of missed spots, we pooled prehybridization LF spot intensities and posthybridization fluorescence spot intensities from all 40 replicates and the pooled data from two cases are shown in Figure 3. In such situations where all local replicates are affected, the prehybrization image allows us to identify and isolate the problem.

Previous reports have shown that variability in prehybridization spot-level intensity, size, and shape irregularities, interrogated using fluorescence methods, influences the accuracy of posthybridization expression measurements.[2,6,13] On the basis of these observations, a quality measure for every spot was developed and filtering based on this quality score was shown to improve consistency in replicate hybridization pairs. We similarly sought to evaluate the impact of spot variability, as measured in our LF images, on our one-color posthybridization fluorescence measurements, by adapting this composite quality scoring method.

Spot quality was evaluated on the basis of three parameters as explained briefly. The measure q_{size} (spot size) assesses spot size irregularity. Spots larger than the usual size are penalized, as this may indicate the presence of contaminants near the spot or that the spot is very close to its neighbors while smaller spots can imply high variability in local background. Size irregularity can also imply that the printing and/or slide conditions were not optimal. This measure is defined as

$$q_{size} = \exp\left(-\frac{|S_{pixel} - S_{pixel0}|}{S_{pixel0}}\right) \quad (4)$$

where S_{pixel} = number of pixels per spot and S_{pixel0} = average number of pixels per spot (computed for each array). $q_{sig\text{-}bkg}$ (spot signal intensity) is a measure of spot density as defined previously in eq 2. q_{sat} (spot saturation) is a measure of pixel saturation in a spot and is indicative of the presence of contaminants that result in a strong intensity value. This measure is defined as

$$q_{sat} = 0 \quad \text{if} \quad S_{sat} \geq 10; \quad \text{else} \quad q_{sat} = 1 \quad (5)$$

where S_{sat} = percentage of saturated pixels in a spot.

Each quality measure is a value between 0 and 1, and the combined score, $q_{lf\text{-}com}$, is defined as

$$q_{lf-com} = (q_{size} \times q_{sig-bkg})^{0.5} \times q_{sat} \quad (6)$$

Figure 4a,b shows prehybridization LF images, and Figure 4c,d shows posthybridization fluorescence images of two subarrays from a slide. Examples of spots with high and low $q_{lf\text{-}com}$ scores are indicated in the images.

We selected a total of four replicate microarray slides, two from each print batch, to evaluate the effect of filtering based on the spot $q_{lf\text{-}com}$ measure. The evaluation was made by monitoring changes in the pairwise correlation coefficient in mean fluorescence intensity measurements of replicate spots from replicate arrays. The mean fluorescence intensity measurements used here were not normalized or corrected for background and represent raw values. The correlation coefficients of the unfiltered data ranged from 0.57 to 0.96. We

(13) Hessner, M. J.; Wang, X.; Khan, S.; Meyer, L.; Schlicht, M.; Tackes, J.; Datta, M. W.; Jacob, H. J.; Ghosh, S. *Nucleic Acids Res.* **2003**, *31*, e60.

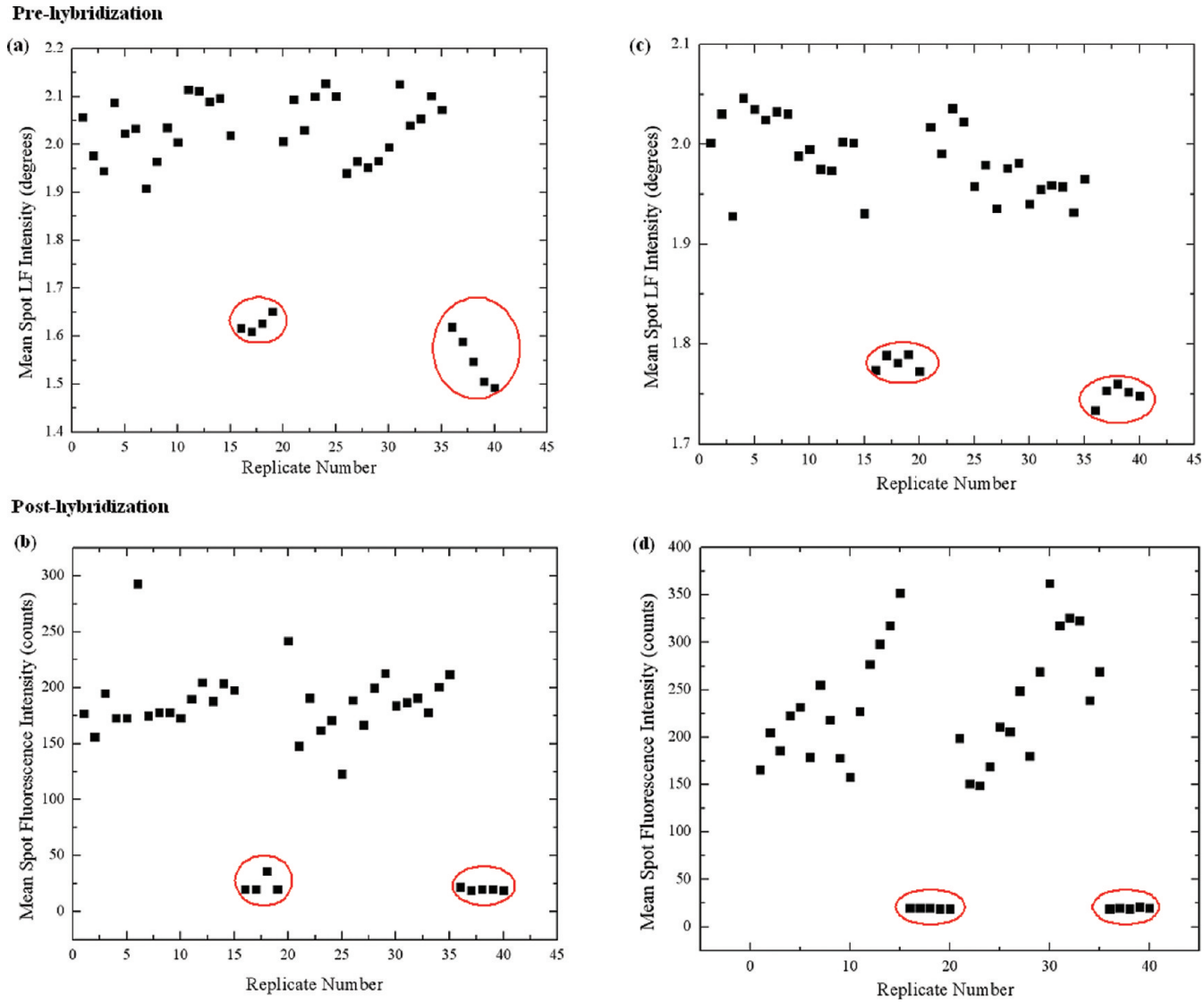


Figure 3. Mean prehybridization spot intensity of 40 replicate spots is shown in (a). Nine of these replicates were missed spots as indicated by the red circles. This finding was corroborated by the posthybridization fluorescence data (b) of the replicates where the same nine spots showed low intensity as indicated by the red circles. Results from another observed instance where 10 of 40 replicate spots were missed are shown in (c) and (d).

found that, by modulating the $q_{\text{lf-com}}$ threshold value, the correlation between replicate arrays was increased to a range of 0.65−0.96 (Table 1). These results indicate that lower correlation between replicates posthybridization is caused in part by prehybridization spot quality variability and that quality filtering using LF images improves reproducibility between arrays.

These quality scores can be utilized to define weights for quality-weighted gene expression analysis where the filtering of poor quality data is achieved through their diminished weights, with possible gains in measurement accuracy.[14]

In defining our spot cumulative quality score $q_{\text{lf-com}}$, we did not include two additional measures presented previously[10] that penalize spots with high variability in local background variability and spots with excessive high local background. This is because spatial variations in the resonant angle due to device imperfections and measurable surface chemistry gradients currently hamper our ability to make meaningful measurements of background variability that reflect the quality of a spot. To address this, the implementation of image registration software that will align and subtract LF images of the device before spotting from images after spotting to remove these nonstochastic sources of background variability is underway.

CONCLUSION

In this work, we have utilized PC biosensors and a high resolution LF imaging detection instrument to generate prehybridization images of spotted DNA microarrays while following standard microarray protocols. Spot level intensity, size, pixel saturation, and local background information was obtained from these images and used for semiautomated microarray quality control. Missed spots from the printing process, which generate false negatives in microarray data, were identified on the basis of spot intensity scores with high accuracy (>95%) and excellent specificity (>98%). Nondiscrete quality scores were generated for the remaining spots based on spot intensity, size, and degree of saturation. We found that variability among replicates was reduced

(14) Gao, S.; Jia, S.; Hessner, M.; Wang, X. *J. Comput. Sci. Syst. Biol.* **2008**; 041−049.

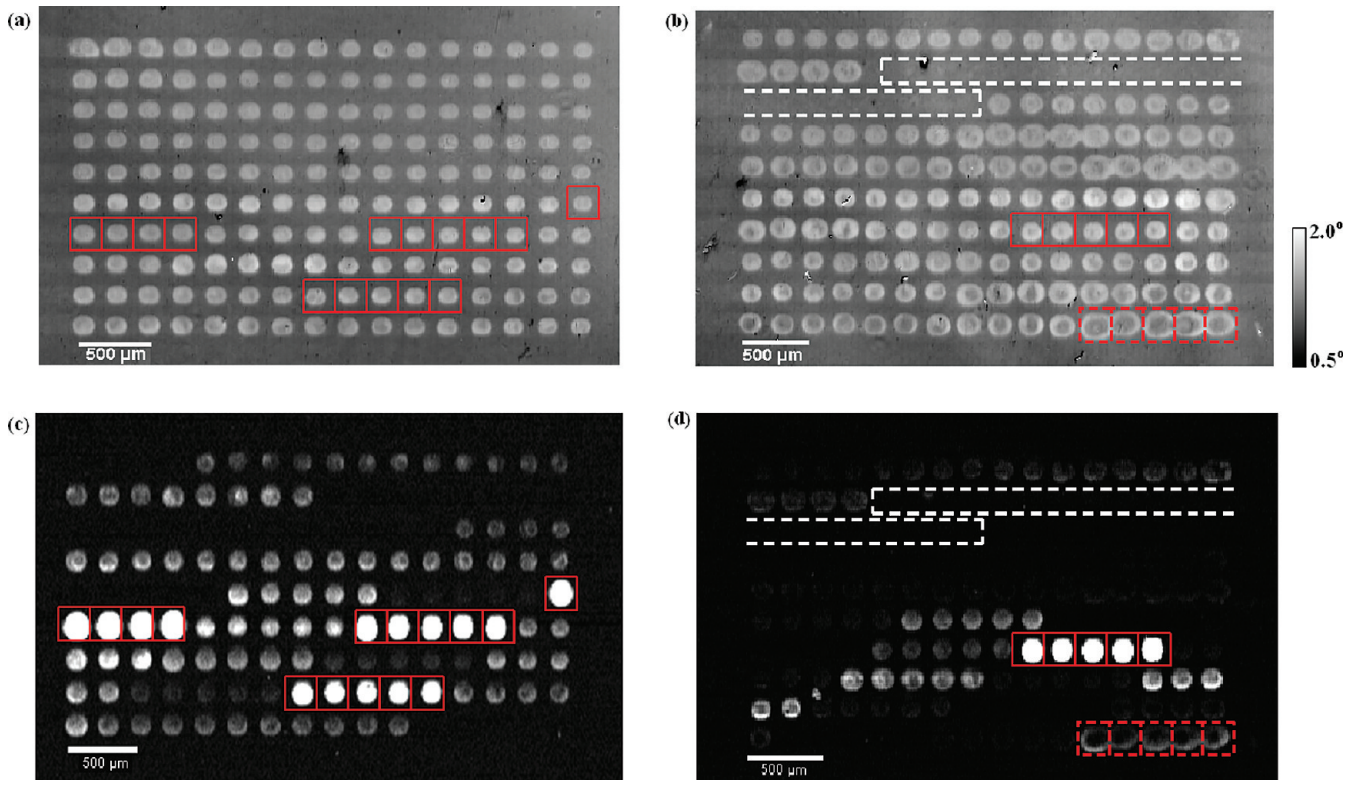


Figure 4. Prehybridization label-free images (a,b) and posthybridization fluorescent images (c,d) of two subarrays from a given slide with spots of differing quality. The white dashed box in (b) and (d) denotes the location of a set of negative control (blank) spots. Open red squares denote spots with $q_{lf\text{-}com} > 0.75$ while open dashed red squares denote spots with $q_{lf\text{-}com} < 0.5$.

Table 1. Correlation Coefficients (r) from the Pairwise Comparisons of Four Replicate Arrays[a]

			r (filtered)		
			$q_{lf\text{-}com}$ threshold		
array no.	array no.	r (unfiltered)	0.6	0.65	0.7
1	2	0.95	0.95	0.96	0.96
1	3	0.72	0.75	0.82	0.84
2	3	0.81	0.81	0.82	0.83
1	4	0.57	0.58	0.64	0.65
2	4	0.89	0.89	0.89	0.92
3	4	0.96	0.96	0.96	0.96

[a] Arrays 1 and 2 are from the first print batch while arrays 3 and 4 are from the second print batch. Coefficient r for the unfiltered raw data is presented alongside data filtered on the basis of $q_{lf\text{-}com}$ scores. Three filtering thresholds were used, and in each, iteration spots below the threshold were eliminated. The $q_{lf\text{-}com}$ score filter eliminated 14−28% of the spots when the threshold was set to 0.65 and 28−66% of the spots with a threshold of 0.7. All values were significant at $p < 0.001$.

by filtering on the basis of quality scores, as demonstrated by higher correlation coefficients. These scores can be utilized as weights in subsequent analysis procedures. Such efforts to identify and reduce the variability introduced prior to hybridization by technical contributors on a slide-by-slide basis will allow microarray users to make more accurate measurements of the more important biological variability that is being studied.

ACKNOWLEDGMENT

This work was supported by grants from the National Institutes of Health (GM086382A), the National Science Foundation (CBET 07-54122), and SRU Biosystems, as well the Linda Su-Nan Chang Sah Doctoral Fellowship (awarded to S.G.). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views National Institutes of Health or the National Science Foundation. The authors thank the staff at the Micro and Nanotechnology Laboratory and colleagues from the Nano Sensors Group at the University of Illinois at Urbana−Champaign.

SUPPORTING INFORMATION AVAILABLE

Additional information noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review June 10, 2010. Accepted August 27, 2010.

AC101551C