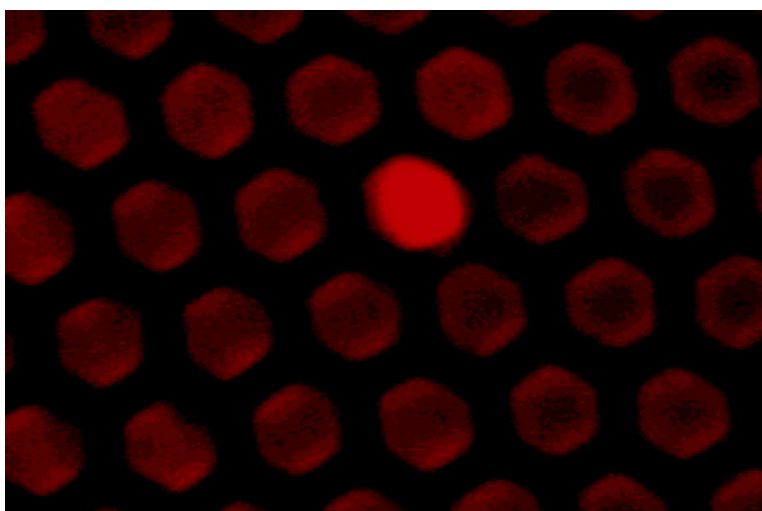## Article

Combinatorial Decoding: An Approach for Universal DNA Array Fabrication

Jason R. Epstein, Jane A. Ferguson, Kyong-Hoon Lee, and David R. Walt

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 6 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Combinatorial Decoding: An Approach for Universal DNA Array Fabrication

Jason R. Epstein, Jane A. Ferguson, Kyong-Hoon Lee, and David R. Walt*

Contribution from the Max Tishler Laboratory for Organic Chemistry, Department of Chemistry, Tufts University, Medford, Massachusetts 02155

Received June 7, 2003; E-mail: david.walt@tufts.edu

Abstract: A fiber optic microsphere-based oligonucleotide array is described that employs the sequence of the oligonucleotide probe attached to each microsphere as positional identifiers. Each microsphere serves as an immobilized array feature, functionalized with a unique single-stranded oligonucleotide sequence and randomly distributed into an array of microwells. To determine the sequences attached to individual microspheres, a series of fluorescently labeled combinatorial-pooled oligonucleotide target solutions was designed. Each combinatorial decoding solution is intended to identify the nucleotide at a particular position on every microsphere in the array. The combinatorial target solutions were synthesized by linking the four possible nucleotides at each position to four different fluorescent reporter dyes. As such, when the solutions were hybridized to the array, one of four possible fluorescent responses was generated for each position on a microsphere probe sequence. Adjusting the stringency of hybridization enabled single-base mismatch discrimination, and the signal with the highest intensity corresponded to the perfect nucleotide match. By consecutively exposing the array to a series of combinatorial decoding pool solutions, it was possible to simultaneously determine the sequence of every randomly positioned oligonucleotide-functionalized microsphere in the array. Once mapped, the microsphere array can be used for any typical genomic microarray experiment.

Introduction

In the past decade, a multitude of novel approaches for genomic analysis has been developed for deciphering the complex nature of genomic information. Of these approaches, microarrays have emerged as likely frontrunner platforms for large-scale, high-throughput, parallel analyses.[1] While microarrays have proven extremely adept at parallel analyses, they have also served to underscore the hurdles inherent with the sheer size of complete genomes. Many microarray formats are currently used for genomic comparisons,[2,3] or single nucleotide polymorphism (SNP) detection,[4,5] but are limited by their feature size and scalability. Scalability is also an extremely important factor in genomic sequencing-by-hybridization,[6,7] where ∼65 000 standard array features are needed to sequence only a few hundred bases. Present methods for large-scale genomic analysis are still restricted either by the overall microarray size or by a particular bias about which sequences are displayed on the array, due to the limited number of features that can be present on the microarray substrate. With the large number of projected human protein-coding genes ($\sim 3 \times 10^4$), human polymorphisms ($>4 \times 10^6$), and the number of genomes still to be sequenced, there are few methods available to correlate complex patterns of gene profiles *in their entirety*. An important remaining challenge is to design an effective analytical method that provides a rapid, accurate, sensitive, and cost-effective approach for large-scale studies. We have previously described a fiber optic platform with the highest packing density of any array format ($\sim 2 \times 10^7$ microsphere sensors/cm^2) that can better address these limitations.[8-10]

The fiber optic microarray platform (Figure 1) entails random self-assembly of oligonucleotide-functionalized microspheres into a fiber optic microwell array.[8-10] The microwell array is formed via selective etching of an imaging fiber bundle.[11] This microsphere-based array platform allows rapid response, high-throughput, extremely low detection limits, and is reusable. A single microsphere-based array has been used more than 100 times with minimal signal degradation.[9] The ability to reuse

(1) Lockhart, D. J.; Winzeler, E. A. *Nature* **2000**, *405*, 827−836.
(2) Tettelin, H.; Masignani, V.; Cieslewicz, M. J.; Eisen, J. A.; Peterson, S.; Wessels, M. R.; Paulsen, I. T.; Nelson, K. E.; Margarit, I.; Read, T. D.; Madoff, L. C.; Wolf, A. M.; Beanan, M. J.; Brinkac, L. M.; Daugherty, S. C.; DeBoy, R. T.; Durkin, A. S.; Kolonay, J. F.; Madupu, R.; Lewis, M. R.; Radune, D.; Fedorova, N. B.; Scanlan, D.; Khouri, H.; Mulligan, S.; Carty, H. A.; Cline, R. T.; Van Aken, S. E.; Gill, J.; Scarselli, M.; Mora, M.; Iacobini, E. T.; Brettoni, C.; Galli, G.; Mariani, M.; Vegni, F.; Maione, D.; Rinaudo, D.; Rappuoli, R.; Telford, J. L.; Kasper, D. L.; Grandi, G.; Fraser, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12391−12396.
(3) Makrigiorgos, G. M.; Chakrabarti, S.; Zhang, Y.; Kaur, M.; Price, B. D. *Nat. Biotechnol.* **2002**, *20*, 936−939.
(4) Lindroos, K.; Sigurdsson, S.; Johansson, K.; Ronnblom, L.; Syvanen, A.-C. *Nucleic Acids Res.* **2002**, *30*, e70/71−e70/79.
(5) Huber, M.; Mundlein, A.; Dornstauder, E.; Schneeberger, C.; Tempfer, C. B.; Mueller, M. W.; Schmidt, W. M. *Anal. Biochem.* **2002**, *303*, 25−33.
(6) Parinov, S.; Barsky, V.; Yershov, G.; Kirillov, E.; Timofeev, E.; Belgovskiy, A.; Mirzabekov, A. *Nucleic Acids Res.* **1996**, *24*, 2998−3004.
(7) Chechetkin, V. R.; Turygin, A. Y.; Proudnikov, D. Y.; Prokopenko, D. V.; Kirillov, E. V.; Mirzabekov, A. D. *J. Biomol. Struct. Dyn.* **2000**, *18*, 83−101.
(8) Epstein, J. R.; Lee, M.; Walt, D. R. *Anal. Chem.* **2002**, *74*, 1836−1840.
(9) Ferguson, J. A.; Steemers, F. J.; Walt, D. R. *Anal. Chem.* **2000**, *72*, 5618−5624.
(10) Walt, D. R. *Science* **2000**, *287*, 451−452.
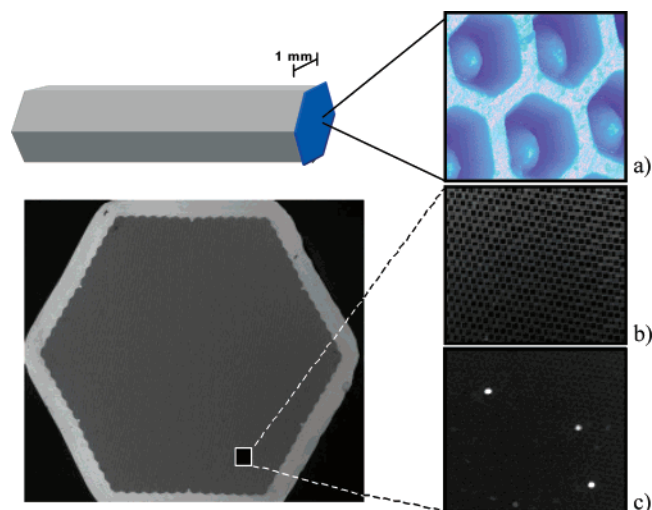(11) Pantano, P.; Walt, D. R. *Chem. Mater.* **1996**, *8*, 2832−2835.

Figure 1. General overview of the fiber optic array platform. (a) The 1-mm diam, hexagonally packed optical fiber bundle is comprised of ~50 000 individual 3.1 μm diam fibers. The fibers are etched, and bead sensors are added to the etched fiber face. (b) A white light image with no fluorescent targets in the array. (c) An image with fluorescent targets hybridized to the array.

Figure 2. General decoding concept. (a) The 20-mer attached to microspheres in the array, consisting of a fixed 8-mer anchor segment and the 12-mer tail segment. The 12-mer tail segment can be any sequence used for genomic analysis; its position in the array is determined by hybridization to combinatorial libraries. (b) The 20-mer decoding solution is comprised of the 8-mer anchor complement and a 12-mer combinatorial tail. The 8-mer anchor ensures correct decoding sequence alignment, and the 12-mer combinatorial tail decodes the sequence on the microsphere. The tail segment length defines the total number of different sequences in the decoding pool ($4^{12} = 1.68 \times 10^7$).

arrays has allowed us to develop a new approach for array decoding and microsphere probe sequence registration. Since the microspheres are randomly distributed into the microwell array, identification of each microsphere location after array assembly is necessary. Currently, two methods exist for determining the sequence on self-assembling microsphere arrays. The first method requires fluorescent dye encoding of microspheres, where each microsphere is labeled with a unique ratio of dyes to identify the attached probe sequence.[9,12]

The use of fluorescent dyes for microsphere encoding is limited by the number of unique, distinguishable optical signatures that can be prepared.[13] The second sequence decoding method employs address sequences on each microsphere and determines the microsphere positions by hybridization to a series of fluorescent complements.[14] The method presented herein eliminates the need to spectrally encode the microspheres or incorporate an address and instead determines the microsphere identity directly via its surface-bound oligonucleotide sequence. In addition to its utility for decoding random probe positions, this method offers the potential for preparing universal arrays with unprecedented feature numbers (every possible 12-mer, ca. 16.7 million sequences).

Combinatorial oligonucleotide libraries were fabricated to determine the sequences on the microspheres in the array. Microsphere sequence identities were determined in parallel by consecutive hybridizations with fluorescently labeled combinatorial nucleic acid decoding solutions. A set of four hybridizations determines the identity of the nucleotide at a given position of every probe sequence in the array. By performing *n* sets of four hybridizations, the sequence of every possible *n*-mer in an array can be differentiated (Figures 2 and 3).

A protocol capable of decoding immobilized 12-mer segments, corresponding to $4^{12} = 1.68 \times 10^7$ unique sequences,
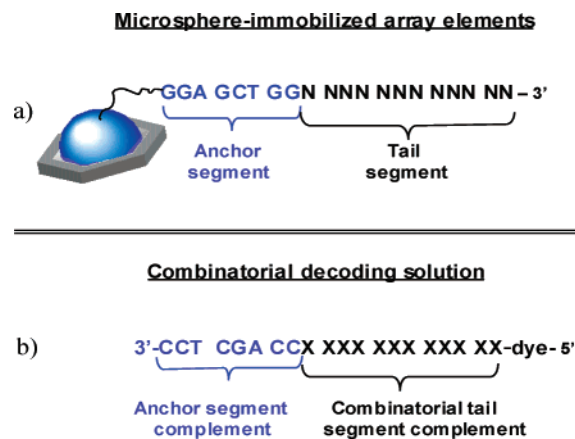
was designed because of its relevance to the coding region of the human genome (~3 × 10^7 bases). This decoding scheme allows the determination of any sequence attached to a microsphere, including unknown, random, or combinatorial sequences.

Materials and Methods

Microsensor Generation. Oligonucleotides were fabricated via standard solid-phase phosphoramidite synthesis methods at the Tufts University Physiology Department (Boston, MA). DNA concentrations were determined by measuring a solution's optical density at 260 nm. DNA sequences were coupled to microspheres as previously described.[8,9] The amine-functionalized microspheres (Bang's Laboratories, Fishers, IN) were encoded by internal trapping of europium (III) thenoyl-trifluoroacetonate·3H$_2$O (Acros) into the microsphere's polymer matrix, prior to oligonucleotide probe attachment. Probe sequences were fabricated with 5′ amine functionality (5′-amino-C6 modifier, Glen Research, Sterling, VA) and were activated for microsphere coupling via cyanuric chloride. The microspheres were stored in PBS buffer containing 0.01% Tween.

Combinatorial Decoding Target Solutions. Target solutions were comprised of the 5′-combinatorial 12-mer followed by the anchor 8-mer complement (Figure 3). For each decoding solution, one position of the 12-mer had a single nucleotide position linked to a 5′-terminal indicator. Regardless of which position on the variable tail segment was fixed, the dye labels linked to each base remained the same: Cy5, Cy3, fluorescein, and a biotin group were linked with determining G, C, A, and T, respectively. Each hybridization required a signal image (post-hybridization) and background image (pre-hybridization), that were subtracted to provide a signal difference. To determine the hybridization extent for T, which incorporated a biotin group, a subsequent hybridization (3 s) with 10 μL of 3 μg/mL streptavidin-fluorescein isothiocyanate was used. Target molecules were diluted in chaotropic salt buffers that leveled the stability differences between G−C and A−T pairs in the single-base discrimination.[15,16]

Array Fabrication. The imaging fiber bundles used for the experiments were obtained from Galileo Electro-Optics Corp. (Stur-

(12) Michael, K. L.; Taylor, L. C.; Schultz, S. L.; Walt, D. R. *Anal. Chem.* **1998**, *70*, 1242−1248.
(13) Braeckmans, K.; De Smedt, S. C.; Leblans, M.; Pauwels, R.; Demeester, J. *Nat. Rev. Drug Discovery* **2002**, *1*, 447−456.
(14) Yeakley, J. M.; Fan, J.-B.; Doucet, D.; Luo, L.; Wickham, E.; Ye, Z.; Chee, M. S.; Fu, X.-D. *Nat. Biotechnol.* **2002**, *20*, 353−358.

(15) Melchior, W. B., Jr.; Von Hippel, P. H. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 298−302.
(16) Wood, W. I.; Gitschier, J.; Lasky, L. A.; Lawn, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1585−1588.
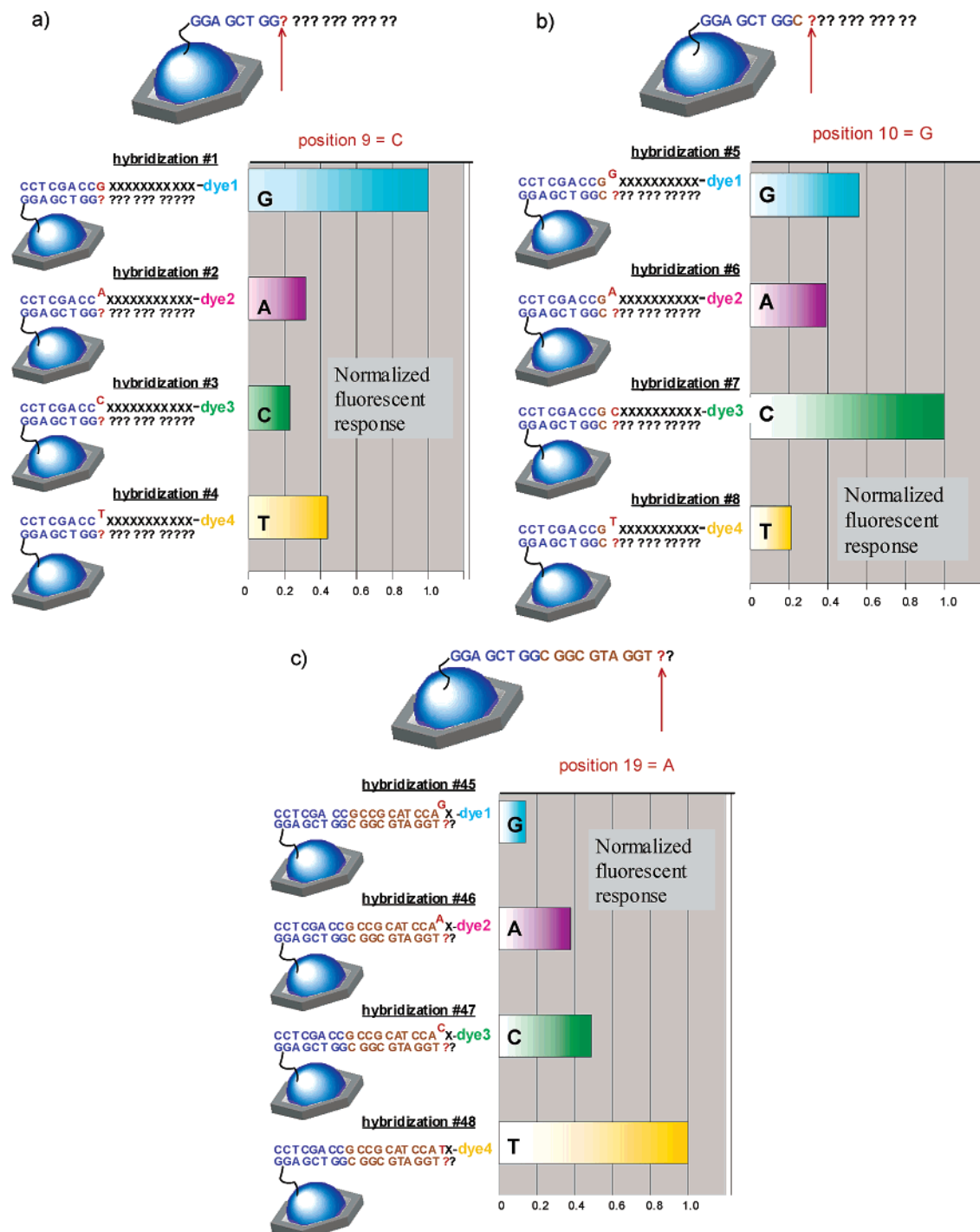
Figure 3. Determination of a sequence on a single bead in the array via consecutive hybridizations. The sequence on the bead consists of the 8-mer anchor and an unknown 12-mer segment (represented by question marks). These figures illustrate the concept that the identity of the base at the first position (position 9 in 3a) can be determined with four hybridizations to decoding solutions. After eight hybridizations, the first two positions (positions 9 and 10 in 3b) can be determined. The process continues the length of the unknown segment to determine every position (Figure 3c). In practice, *because each decoding solution acts on the entire array, the nucleotide identity at a given position on every bead in the array is determined simultaneously.*

bridge, MA) and Illumina Inc. (San Diego, CA). The fiber was polished and etched as previously described.[8,9] Aliquots of the microspheres were added to the etched fiber face sequentially, and their position in the array was determined by the encoding dye.

Imaging System. The custom-built imaging systems were modified Olympus microscopes (Melville, NY), with charge-coupled device (CCD) cameras, either Hamamatsu Orca ER (Bridgewater, NJ) or Princeton Instruments (intensified CCD, Trenton, NJ). Images were collected and analyzed using IPLab software (Scanalytics, Fairfax, VA). For each experiment, the results from each microsphere were normalized to the highest signal of the four different dyes. The camera exposure times were set to produce comparable signals from the different dyes, accounting for differences in quantum and labeling efficiency. Camera exposure times were 100 ms for Cy3 and Cy5, and 2 s for the fluorescein measurements.

Decoding Experiments. Decoding target pools are listed for the base and position they were used to decode. For example, decoding target solution 9A was used to label all microspheres containing an A at position 9. The optical density (ODU) for each decoding target was determined and stored in 1 mL SBB buffer (0.1 M sodium borate buffer,

pH 8.3) to form the decoding stock solutions. Any further dilutions were performed with 2.4 M TEA buffer (2.4 M tetraethylammonium chloride, 50 mM TRIS, 5 mM EDTA, pH 8.0, filtered 0.22 μm) and formamide (F). Hybridizations were performed with 20 μL of target solution except for 9A, 15A, and 19A, which were hybridized in a total volume of 60 μL. In addition, all Cy3 signals and streptavidin-fluorescein signals were decreased by a factor of 4 and 2, respectively, via mathematical scaling. **Position-9** series of four hybridizations were performed with 20-min hybridizations. Decoding solution 9A (21.02 ODU) was diluted 1:2 and contained 20%F, with a total volume of 60 μL. Decoding solution 9T (23.28 ODU) was diluted 1:2 and contained 20%F. Decoding solution 9G (21.48 ODU) was diluted 1:5 and contained 20%F. Decoding solution 9C (21.36 ODU) was diluted 1:5 and contained 20%F. **Position-15** series of four hybridizations were performed with 20-min hybridizations. Decoding solution 15A (28.82 ODU) was diluted 1:2 and contained 5%F, with a total volume of 60 μL. Decoding solution 15T (34.08 ODU) was diluted 1:5 and contained 20%F. Decoding solution 15G (20.28 ODU) contained no F. Decoding solution 15C (20.28 ODU) was diluted 1:5 and contained 20%F. **Position-19** series of four hybridizations were performed with 45-min hybridizations. Decoding solution 19A (24.60 ODU) was diluted 1:2 and contained no F, with a total volume of 60 μL. Decoding solution 19T (24.36 ODU) was diluted 1:5 and contained 20%F. Decoding solution 19G (21.98 ODU) contained no F. Decoding solution 19C (21.79 ODU) was diluted 1:5 and contained 20%F.

Results

Multiple cycles of hybridization and dehybridization were used to directly sequence the oligonucleotide attached to the microsphere surface. While numerous array platforms have been used previously to sequence genomic samples,[17,18] in the present case we are actually determining the sequences immobilized to the microspheres in the array. By "sequencing the array", we can simultaneously determine the sequence of every ssDNA sequence element present.

Every microsphere contained a 20-mer oligonucleotide sequence comprised of a known 5′ 8-mer "anchor sequence" followed by a 12-mer "tail" segment (Figure 2a). The decoding pool libraries were also 20-mer, comprised of the 3′ anchor-complement, a 12-mer combinatorial segment, and a 5′-fluorescent dye (Figure 2b). The 8-mer anchor was uniform for every microsphere and was the perfect complement to the 3′ 8-mer portion of the decoding pool. The 8-mer anchors ensured that the decoding pool 20-mer sequence complements, when hybridized, were correctly aligned to the 20-mer microsphere sequence. The difference between each microsphere is within the 12-mer tail segment of the surface-bound oligonucleotide (Figure 4). The target decoding pools were fabricated via standard solid-phase phosphoramidite chemistry, using combinatorial synthesis. Hybridization of the pools to the array determined the identities of the different 12-mer tails (conceptually illustrated in Figure 3). The combinatorial synthesis strategy of each decoding pool ensured that there would be a perfect complement to every 12-mer tail segment. For example, four combinatorial sub-libraries were fabricated to determine the nucleotide at the first position of an unknown 12-mer tail segment. The four decoding sub-libraries were identical except for the nucleotide at the first position (nucleotide 9 in the 20-mer) and the dyes attached to their 5′ termini, such that library 9G had a guanosine

(17) Strezoska, Z.; Paunesku, T.; Radosavljevic, D.; Labat, I.; Drmanac, R.; Crkvenjakov, R. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10089−10093.
(18) Drmanac, S.; Kita, D.; Labat, I.; Hauser, B.; Schmidt, C.; Burczak, J. D.; Drmanac, R. *Nat. Biotechnol.* **1998**, *16*, 54−58.

Figure 4. The 12 sequences attached to microspheres and decoded in the array. The 12-mer sequences were based on the K-ras wild-type sequence, an oncogene implicated in the development of human malignancies. The 12 sequences were chosen to bracket in the entire 11-mer segment from position 9 to position 19.

at the first position, linked to a 5′-Cy5 dye, library 9C had a cytidine at the first position, terminated with a 5′-Cy3 dye, etc. Thus, the four sub-libraries in toto contain every possible 12-mer tail sequence, but individually, each sub-library contains only one of the four nucleotides at position 9 linked to a unique fluorescent dye. Specifically, for decoding the identity of the base at position 9, each sub-library contains a pool of every possible nucleotide combination at the 11 unspecified positions (positions 10−20, 4.2×10^6 total sequences per pool). After each round of four hybridizations, the array is dehybridized to allow a second set of four hybridizations. A second set of four sub-libraries is hybridized to generate another set of four-color responses to identify every nucleotide at the second position (nucleotide 10 in the 20-mer). This process continues for *n* rounds of four hybridizations, for a sequence of length *n*, enabling complete sequence identification for every microsphere in the array.

Hybridizing each of the four libraries to the array separately under single-base mismatch stringency conditions enables only the perfect sequence complement, or the closest sequence match in that pool, to hybridize to a given microsphere (Figure 5). Only one sub-library pool contains the perfect complement, while the other three can, at best, yield single-base mismatches. By hybridizing to the four pools, the library containing the perfect complement provides the highest signal response, enabling a determination of the correct nucleotide identity at that position of the 12-mer tail. In this paper, we demonstrate how this approach was used to determine the identity of specific nucleotides in the array at the 9th, 15th, and 19th positions of the 20-mer, in principle allowing the sequencing of 11 consecutive tail bases of every microsphere-immobilized sequence in the array (Note: position 20 was not attempted).

Decoding of a 12-mer Tail Segment. To demonstrate that the oligonucleotides on the microspheres could be sequenced in a multiplexed fashion, 12 different microsphere types were employed for this study. Because there is no independent way to identify the probe sequences in a more complex microsphere library comprised of hundreds of different bead types, we chose to limit ourselves to the 12 different microspheres. The 12
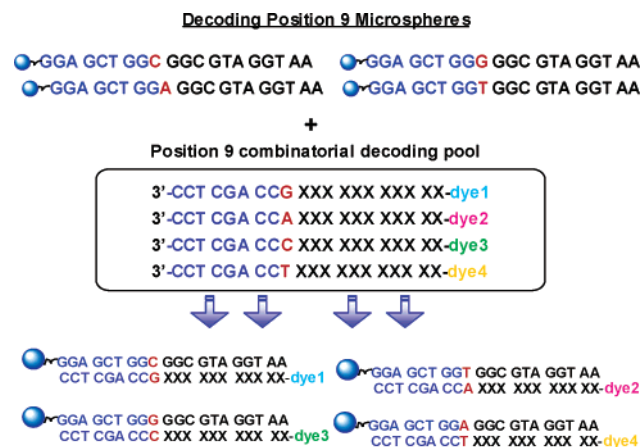
Decoding Position 9 Microspheres

Figure 5. Four exemplar microspheres used for position 9 determinations. The four possible position 9 nucleotides were determined via hybridization to the four combinatorial sub-library pools. Each position 9 nucleotide has a perfect match in only one of the four sub-libraries. The remaining 11 nucleotides have a perfect complement in each of the four combinatorial libraries. "X" represents equal portions of A, G, C, and T.

different microsphere types were chosen to represent all four nucleotides at each of three different positions on the 12-mer tail (Figure 4). Each microsphere type was embedded with an encoding dye to track its random positioning in the array and to allow us to determine the fidelity of the sequencing. The three positions (9th, 15th, and 19th from the 5′ end) were chosen to demonstrate the ability to determine any sequence within the first 11 tail positions including the ends (positions 9 and 19) and the middle (position 15). Microspheres were fabricated in large batches; therefore, multiple arrays with identical microspheres could be produced. Once an array was fabricated, the fluorescent encoding dye was used to associate each microsphere with its position in the array. Multiplexed arrays were subjected to multiple rounds of hybridization to determine the sequence of the three positions on the 12-mer tail. To sequence a single position on the 12-mer tail, four consecutive hybridizations were performed, providing four signal responses corresponding to each nucleotide (Figure 2). Each encoding dye was linked to a known sequence, thereby confirming the experimentally determined sequencing efforts.

Table 1 shows the results from sequencing the 9th, 15th, and 19th positions of the 20-mer oligonucleotide sequences in the array. The results from sequencing the 9th position are also represented in Figure 6. Because the array fabrication process is random, each array was represented with multiple copies of identical microsphere types. For example, one array used in decoding the 9th position contained a total of 25 microspheres (11, 9C; 8, 9G; 3, 9A; and 3, 9T). While the results in Table 1 show the average responses of every identical microsphere in the array, *every nucleotide on every microsphere at position 9 was correctly identified*. In some instances, the nonspecific signal of a single bead was ∼95% that of the correct match, but because the experiment is designed as a "winner takes all" approach, the correct identification was always possible. Performing the decoding over multiple arrays, representing hundreds of microspheres, still yielded >99% correct nucleotide identification. As decoding proceeded from position 9 toward more terminal positions, the identity of the base became more difficult to distinguish as would be expected. The higher nonspecific signals associated with decoding positions 15 and

Table 1. Responses for Sequential Decoding of Random Microspheres in the Array[a]

		C	G	A	T	
position 9	C	1.00	0.153	0.312	0.355	all correct
	G	0.326	1.000	0.200	0.294	all correct
	A	0.660	0.372	1.000	0.819	all correct
	T	0.680	0.111	0.753	1.000	all correct
position 15	C	1.000	0.434	0.208	0.870	all correct
	G	0.546	1.000	0.701	0.810	all correct
	A	0.725	0.437	1.000	0.683	all correct
	T	0.530	0.298	0.523	1.000	all correct
position 19	C	1.000	0.367	0.306	0.733	all correct
	G	0.544	1.000	0.210	0.693	all correct
	A	0.750	0.307	1.000	0.695	all correct
	T	0.655	0.370	0.443	1.000	all correct

[a] Results presented are from three different arrays, one for the 9th, 15th, and 19th positions. All values were normalized to the highest signal for each microsphere so the intended target responses equaled 1.000 (shown on the diagonal) and are presented as averages of idential microspheres. In each case, every individual microsphere provided the correct response corresponding to the decoded position. The number of beads in each assay was as follows: (a) position 9 microspheres 9C (11), 9G (8), 9A (3), 9T (3); (b) position 15 microspheres 15C (2), 15G (3), 15A (6), 15T (14); and (c) position 19 microspheres 19C (11), 19G (3), 19A(5), 19T (11). The results from sequencing position 9 are also represented graphically in Figure 6.
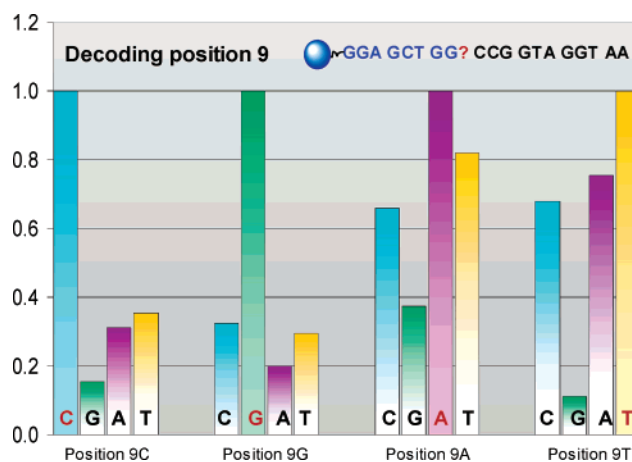

Figure 6. Experimental results from decoding position 9 of a 12-mer tail segment. Position 9 of every bead in the array, including all four different microsphere types (9C, 9G, 9A, and 9T), were determined via four hybridizations to the array. The results were normalized to the highest signal for each microsphere, and identical microspheres are presented as average values.

19 are seen in Table 1. Although similar levels of success were achieved at these positions, the results from both positions have an increase in nonspecific signal, and the experiments decoding position 19 required extending the hybridization time twice that necessary for successful decoding of position 15.

Decoding Specific Arrays. While decoding unknown or universal sequences in the array would require full-length sequential decoding, an array containing a restricted number of sequences is substantially less complex. In such arrays, the microsphere sequence positions to be decoded can be selected to optimize the decoding libraries' differentiating capability. For example, an array containing ∼4×10^3 probe sequences (of any length) could be completely determined with six rounds of decoding *at any of six positions* (four nucleotides and six decoding pools are capable of decoding $4^6 = 4096$ sequences). Successful decoding of eight tail segment nucleotides would enable discrimination of more than 65 000 (4^8) different array features. At this capacity, the largest arrays currently employed

for incorporating specific sequences could be unambiguously decoded.

This simplified decoding concept was demonstrated by fabricating an array with different known tail segments. Different 12-mer tail segments were chosen such that, on the basis of determining only the 9th and 15th positions, the identity of the 19th position could be deduced. The 9th and 15th positions were determined with their respective decoding target pools (data not shown). Hybridizations to determine the 19th position were unnecessary because the 9th and 15th positions unequivocally specified the sequences. This concept illustrates that if predetermined sequences are added to the array, extraneous decoding steps can be avoided, and the sequence information can be obtained from the information provided by other bases in the probe.

Discussion

The decoding method described in this paper involves a "winner takes all" approach, where the dominant fluorescence signal defines each nucleotide in the microsphere tail sequence. Because there are $4^{11} = 4.2 \times 10^6$ sequences in every pool, the specific complement is present at a concentration of $\sim$1 pM in a background of $\sim$100 μM, which results in a substantial nonspecific signal. Even so, a 20 μL sample of the combinatorial decoding solution still contains $\sim$10^6 perfect complement molecules, providing ample target numbers to generate a specific signal. The fiber optic platform has previously been used to detect target concentrations 3−4 orders of magnitude below this range.[8] Extending this sequencing method beyond the 12-mer tail to longer variable tail regions should be possible, however, on the same synthetic scale; extending this sequencing protocol to longer sequences would eventually dilute the number of perfect complements to a level below the detection limit.

The differences between the complementary target hybridization signal and the three mismatched hybridization signals were a result of both the labeling dye and the tail segment position interrogated. Three of the sequence pools were linked to different fluorescent phosphoramidite-dyes (Cy3, Cy5, and fluorescein), while the fourth sequence pool was terminated with a phosphoramidite-biotin group, which required an additional streptavidin/fluorescein binding step. Because the protocol incorporates four different indicators linked to decode each nucleotide, the assay offers the potential to determine each position of the tail segment simultaneously. Our initial design was to multiplex the determination of each nucleotide in a single hybridization using all four colors in a single hybridization. By performing four separate hybridizations (a separate hybridization to interrogate each possible base) for each position of the tail segment, we obtained much higher specific signals. The higher signals were likely due to the lower complexity of the hybridization solutions and allowed shorter hybridization times, thereby increasing the throughput of the assay. For example, while the single four color multiplexed pooled assay showed success in simultaneously decoding position 19, the required hybridization time was 12 h (data not shown). By dividing the decoding pools into four separate hybridizations, a single hybridization was completed within 45 min, saving more than 9 h in experimental time. In addition, by dividing the decoding pools into four sub-libraries, we could use the same dye for every nucleotide, eliminating corrections accounting for differences in dye quantum efficiencies. Currently, the multiple dyes

and the extraneous streptavidin/fluorescein hybridization step are being replaced by a single reporter dye.

The fluorescein phosphoramidite label generated significantly lower signals than the fluorescein-streptavidin label. This low fluorescence was found to be unrelated to hybridization efficiency, as the same ratios in signal response were seen when the base/dye combinations were rearranged. To counteract this effect, higher concentrations of hybridization target based on the fluorescein phosphoramidite were used in comparison to the biotinylated fluorescein targets.

The ultimate goal would be to decode every position of every sequence in an array, creating a map of the array based on the identity of the oligonucleotide sequences present. Our experiments were performed with arrays containing only a few dozen 3 μm diameter features, but for large-scale assays, it would be preferable to simultaneously sequence more than 50 000 microspheres on the array. Currently, a single decoding hybridization employs 20−60 μL of decoding pooled target solution, which can be used for multiple arrays. To increase the relative target numbers or the signal response, higher concentrations or volumes may be necessary. Although successful determination of the 19th position was obtained, as expected, the more terminal positions were less successful and required longer hybridization times. This difficulty stems from the complexity of the decoding solution and the lower fidelity of hybridization associated with mismatches at more terminal sequence positions. The lower terminal base success rate may be countered by flanking the variable tail segment with an additional anchor sequence, or by improving the hybridization with a simplified decoding pool that employs universal bases such as nitroindole[19] or deoxyinosine.[20] Finally, while the protocol described here used an anchor sequence to correctly align the decoding 20-mer, in directed sequences, a fully combinatorial decoding solution can be employed, as there is no need to include an alignment sequence.

Conclusions

In this paper, we have demonstrated an initial effort to develop a more versatile microarray platform. The decoding method incorporates the inherent advantages of fiber optic microsphere arrays, including miniature feature sizes, simple fabrication protocols, rapid assay times, and a reusable platform of regenerable sensors. The ability to reuse arrays is a major advantage of the fiber optic platform, as array-to-array variability has been widely documented.[21,22] The protocol described in this manuscript provides an improved method to register the random positions of the microsphere array elements, enabling the use of extremely high-density array platforms for comprehensive genomic applications. A key component of this approach is the parallel and efficient sequence determination, and it is an important first step toward future microsphere-based applications.

This method is only the initial phase of numerous future research directions involving this platform, including genomic sequencing-by-hybridization, SNP detection, and comprehensive

(19) Loakes, D.; Hill, F.; Brown, D. M.; Salisbury, S. A. *J. Mol. Biol.* **1997**, *270*, 426−435.
(20) Cubero, E.; Guimil-Garcia, R.; Luque, F. J.; Eritja, R.; Orozco, M. *Nucleic Acids Res.* **2001**, *29*, 2522−2534.
(21) Lee, M.-L. T.; Kuo, F. C.; Whitmore, G. A.; Sklar, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9834−9839.
(22) Spruill, S. E.; Lu, J.; Weir, B. *BioTechniques* **2002**, *33*, 916−920, 922−923.

diagnostic assays. The platform size allows the highest packing density of any array platform and, as such, is better suited for large-scale, unbiased genomic analysis. This sequencing–decoding method can determine the microsphere probe sequences, but also could serve as a quality control method to avoid costly fabrication errors.[23,24] The microsphere-based platform is expected to be further developed into a *universal*

oligonucleotide microarray format, more applicable to the interrogation of entire genomes.

JA0365577

(23) Knight, J. *Nature* **2001**, *410*, 860–861.
(24) Marshall, E. *Science* **2001**, *291*, 2535.