TECHNICAL NOTE

Jennifer T. Kemp,[1] Ph.D.; Ronald W. Davis,[1] Ph.D.; Robert L. White,[2] Ph.D.;
Shan X. Wang,[2] Ph.D.; and Chris D. Webb,[1] Ph.D.

A Novel Method for STR-based DNA Profiling Using Microarrays

ABSTRACT: We describe a novel method for rapidly identifying and distinguishing between different DNA sequences using short tandem repeat (STR) analysis and DNA microarrays. The method can be used to deduce identity, length, and number of STRs of the target molecule. We refer to this technique as the "variable-length probe array" method for STR profiling (VLPA). The method involves hybridization of the unknown STR target sequence to a DNA microarray displaying complementary probes that vary in length to cover the range of possible STRs. A post-hybridization enzymatic digestion of the DNA hybrids is then used to selectively remove labeled single-stranded regions of DNA from the microarray surface. The number of repeats in the unknown target is then deduced based on the pattern of target DNA that remains hybridized to the array. This DNA profiling technique is useful for performing forensic analysis to uniquely identify individual humans or other species.

KEYWORDS: forensic science, short tandem repeat, DNA profiling, DNA fingerprinting, microarrays, hybridization, S1 nuclease, endonuclease, VLPA

DNA-based techniques for the identification of individuals are becoming increasingly relied upon in forensic science (1). The profiling method (sometimes called "DNA fingerprinting") that the FBI and the British courts have accepted for use in identification of an individual is based on the tandem repeats present in the human genome (2–4). In the noncoding regions of the genome, there are many loci where a particular sequence of DNA is repeated multiple times in direct succession. The number of tandem repeats at a given DNA locus varies between individuals. The loci frequently used in forensic science consist of STRs (short tandem repeats) and typically contain about 3 to 15 repeats, each with between 3 and 7 base pairs. While longer repeats also exist in the genome, the shorter repeats (usually 4 or 5 base pairs) are most often used in forensic analysis, since the short repeat regions are readily amenable to PCR amplification. The FBI and the forensic science community typically use 13 separate STR loci (the core CODIS loci) in routine forensic analysis (1,4). If all 13 loci have identical lengths in two DNA samples, the probability that the two samples originated from different specimens is low enough that the courts generally accept this identification as definitive evidence that the individuals in question are the same (1).

To perform a DNA profiling experiment based on STR analysis, electrophoretic profiles of the regions of DNA corresponding to each of the 13 STR loci are obtained and compared between samples (2). Miniature systems with an array of electrophoretic columns

for this purpose have been developed, and use of the technique is widespread (5–7). The Department of Justice predicts that STR analysis will remain the technique of choice in forensic science for DNA identification for the next decade, and that the number of loci used in this analysis will perhaps be increased from 13 to 20 (1).

While these electrophoretic DNA profiling methods are based on mature technologies, DNA profiling methods using microarrays are in their infancy. Microarray-based assays are desirable since they are compatible with miniaturized devices that could provide a high degree of speed, sensitivity, and portability, which are important features in forensic analysis. Several methods using microarray assays for identification of single nucleotide polymorphisms (SNPs) have been developed (8,9). However, these methods require either special electrically active Nanogen[TM] DNA chips or sophisticated tiling probe sets to identify a single SNP, and have not been widely adopted in forensic analysis.

One method that discriminates between STR alleles on microarrays has been described (10). In principle, a target containing an STR of unknown repeat length can be hybridized to an array displaying complementary probes that vary in length to cover the range of possible number of repeats. Differences in hybridization of target DNA to the various probes can then be used to determine the number of repeats; for example, a target with 10 repeats should bind more strongly to a probe with 10 repeats than to a probe with 5. However, in practice, the difference in hybridization efficiency of tandem repeats that are similar in length (such as 9 and 10 repeats) is very subtle and may be hard to detect. Radtkey and colleagues (10) describe a high stringency approach to discriminate between repeats of similar lengths. However, this requires an electronically active DNA array to allow discrimination of subtle hybridization differences. Here we describe a new method, VLPA (for Variable-Length Probe Array), to determine the length of an unknown STR using two novel technical innovations, a clamp sequence to ensure

TABLE 1—*Oligonucleotides used in the VLPA experiment*.

Oligonucleotide	Function	Repeats	Sequence
JTK026-r	probe	1	[AminoC6] GTACCGGAATTCCGG <u>ACGTGACTCT</u>
JTK027-r	probe	2	[AminoC6] GTACCGGAATTCCGG <u>ACGTGACTCT ACGTGACTCT</u>
JTK028-r	probe	3	[AminoC6] GTACCGGAATTCCGG <u>ACGTGACTCT ACGTGACTCT ACGTGACTCT</u>
JTK028	target	3	[Cy5] <u>AGAGTCACGT AGAGTCACGT AGAGTCACGT</u> CCGGAATTCCGGTAC

Note: Probes consisted of a 5′ amino modification with a C6 spacer, 1, 2, or 3 repeats of the 10 bp sequence ACGTGACTCT (underlined), and a 15 bp clamp sequence (not underlined). The target was labeled with a 5′ Cy5 fluorophore and had 3 repeats of the 10 bp sequence AGAGTCACGT (underlined, complementary to the probe repeat sequence) and a 15 bp clamp sequence (not underlined, complementary to the probe clamp sequence).

proper hybridization of the repeat sequences and a nuclease step to remove single-stranded DNA from the array. This allows us to deduce the number of repeats from the resulting fluorescent signal pattern. This method utilizes currently widely available microarray technology and should allow rapid determination of individual identity.

Materials and Methods

Microarrays were prepared using CodeLink activated slides (Amersham, Piscataway, NJ) and 5′ amine-modified oligonucleotide probes (Qiagen, Alameda, CA). The oligonucleotides (5′–3′) consisted of a 5′ amine group (for attachment to the array), a 15 bp clamp sequence, and 1, 2, or 3 tandem repeats of a 10 bp sequence (Table 1). Probes were printed onto microarrays from a solution containing the oligonucleotide at a concentration of 10 μM using an OmniGrid microarrayer (GeneMachines, Ann Arbor, MI). The post-printing processing of the microarrays was performed as recommended by the slide manufacturer.

Hybridization was performed using a target oligonucleotide (Qiagen, Alameda, CA) consisting of (5′–3′): a Cy5 fluorophore on the 5′ end, three tandem repeats of a 10 bp sequence that was complementary to repeats on the probe, and a 15 bp sequence complementary to the clamp on the probe (Table 1). The target oligonucleotide was applied to the microarray at a concentration of 1 μM and the hybridizations were performed at 50°C for 4–12 h. After hybridization, the microarrays were washed 3 times in SSC buffer (Amersham protocol) at room temperature and then submerged into buffer that was pre-equilibrated to 37°C and that contained S1 endonuclease (Invitrogen, Carlsbad, CA) at 0.3 μL/mL in 1x reaction buffer. Microarrays were then incubated in S1 endonuclease solution at 37°C for 10 min with intermittent agitation. After nuclease digestion, microarrays were washed three times in buffer containing 0.01X SSC and 0.01% SDS, three times in buffer containing 0.01X SSC, and dried. Microarrays were assayed for fluorescent signal at 635 nm using a GenePix 4000 fluorescent scanner (Axon Instruments, Foster City, CA) set to scan at 400 PMT.

The experiments detailed in the present study were performed using a 10 min S1 nuclease incubation, which we determined to be optimal. In other experiments (data not shown), some digestion was apparent after as little as 2 min, while loss of signal due to overdigestion was observed when incubation proceeded 15–30 min or longer. The signal differential between probes was greatest at 10 min.

To quantitate fluorescence intensity from each probe (Table 2), we used GenePix Pro software to determine the total fluorescent signal from each feature. Four separate arrays were analyzed for each treatment and the results were compiled as follows. For each oligonucleotide under each condition, data was collected from at least 6 separate features from the control experiments (hybridization experiment and buffer incubation), and from 14 separate features

TABLE 2—*Mean fluorescence intensities from VLPA feasibility experiment*.

Control 1: After hybridization, no nuclease incubation			
	3-repeat probe	2-repeat probe	1-repeat probe
A	100 ± 10	104 ± 27	103 ± 11
B	100 ± 14	123 ± 13	101 ± 8
C	100 ± 5	121 ± 7	81 ± 4
D	100 ± 3	103 ± 4	89 ± 3
Mean	**100**	**113**	**94**
Control 2: Incubation in nuclease buffer without nuclease			
	3-repeat probe	2-repeat probe	1-repeat probe
A	100 ± 11	117 ± 26	120 ± 12
B	100 ± 11	147 ± 12	101 ± 6
C	100 ± 9	137 ± 15	97 ± 9
D	100 ± 3	127 ± 5	97 ± 5
Mean	**100**	**132**	**104**
Nuclease incubation			
	3-repeat probe	2-repeat probe	1-repeat probe
A	100 ± 5	42 ± 2	7 ± 0.3
B	100 ± 8	71 ± 5	32 ± 1
C	100 ± 4	59 ± 3	22 ± 1
D	100 ± 6	77 ± 6	19 ± 1
Mean	**100**	**62**	**20**

Note: The mean fluorescence intensities (expressed as a percentage of the 3-repeat probe intensity) plus or minus the standard error of the mean (SEM) are indicated for the control arrays and the nuclease test arrays. Data from four independent but identical experiments (A, B, C, and D) are displayed, and the means of all four experiments are displayed in bold.

from each nuclease incubation experiment. The standard error of the mean (SEM) was calculated for each fluorescent dataset (Table 2). Unpaired t-tests were used to calculate p values for the data from the nuclease treatment. In all experiments, background fluorescence was less than 5%.

Results

Strategy to Determine STR Length Using VLPA

Detection of STR length using microarrays is hampered by the fact that the hybridization efficiency of repeats that are close in length is very similar. This makes it hard to distinguish between STRs with similar numbers of repeats. To overcome this problem, we have designed the VLPA method, which is performed as follows. Single-stranded DNA probes with varying number of repeats (and thus variable length) are end-attached to a microarray surface (each probe to a separate feature or "spot") (a hypothetical array with up to five repeats is diagrammed in Fig. 1*A*). Next, a sample containing fluorescently end-labeled single-stranded DNA with an unknown number of STRs is applied to the microarray and allowed to hybridize (a hypothetical target with three repeats is diagrammed in Fig. 1*B*; hybridization of this target to the array is depicted in Fig. 1*C*). After hybridization, the microarray is subjected to
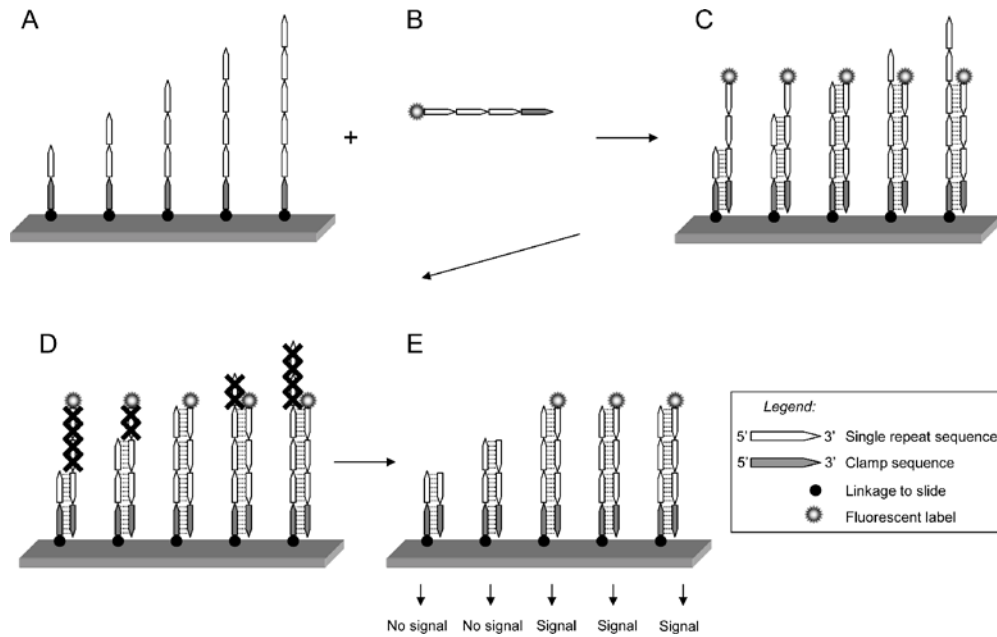
FIG. 1—*Variable-length probe array (VLPA) procedure for STR-based DNA profiling. (A) Probes with 1, 2, 3, 4, and 5 repeats are attached to the surface of the array. (B) The labeled target (unknown) has 3 repeats in this hypothetical example. (C) The target is hybridized to the array. (D) The single-stranded regions that result upon hybridization are digested by the nuclease (indicated by X marks over the digested regions of DNA), and the signal is decreased from features having probes with 1 or 2 repeats. (E) Signal persists in probes with 3, 4, or 5 repeats; therefore the conclusion is that the target is 3 repeats long.*

enzymatic digestion using a single-stranded endonuclease. This treatment removes single-stranded regions of DNA and consequently, removes the fluorescent label from the end of any single stranded region protruding from a hybridized duplex (Fig. 1D).

Three possible outcomes exist for the resulting target-probe hybridization pattern. The first possible outcome is that the labeled target may have more repeats than the probe attached to the microarray. As described below, we use a clamp sequence to ensure that the target DNA anneals to the probe so that the single stranded region of the target DNA will protrude from the hybridized complex into solution (see example of probes with 1 and 2 repeats in Fig. 1C). When the microarray is treated with single-stranded endonuclease, the single-stranded region of target DNA and the fluorescent label are removed (see example of probes with 1 and 2 repeats in Fig. 1D and E), resulting in a loss of signal detected from this feature.

The second possible outcome is that the target and the probe may have an equal number of repeats, in which case no single-stranded DNA is present (see example of probe with 3 repeats in Fig. 1C). In this case, the endonuclease treatment has no effect on the hybridized complex and the fluorescent moiety is not removed (see example of probe with 3 repeats in Fig. 1D and E). The signal detected from this feature remains unchanged.

The third outcome occurs if the target has fewer repeats than the probe, in which case a region of single-stranded probe DNA protrudes from the hybridized complex (see example of probes with 4 and 5 repeats in Fig. 1C). Although this single-stranded region of probe DNA is removed during the endonuclease treatment, the target DNA is not digested and the fluorescent label remains attached (the signal from this feature remains unchanged).

Thus, following endonuclease treatment, the fluorescent signal will only remain on features containing probes with an equal or greater number of repeats than the target (Fig. 1E). The fluorescent signal can now be read using a standard microarray scanner without any additional special equipment. The number of repeats in the unknown target DNA is deduced from the results of the enzymatic digestion of the hybridized microarray and is determined to be equal to the number of repeats of the shortest probe that yields fluorescent signal after enzymatic digestion.

A key requirement for this strategy to work is that the target anneals to the probe in the proper register; that is, it must anneal without misaligned repeats or "slippage." For example, in Fig. 2A, a target with more repeats than the probe could anneal such that the fluorophore would not be removed by nuclease treatment and an improper signal would be retained. Conversely, in Fig. 2C, a target with fewer repeats than the probe could anneal such that the fluorophore would be removed by the nuclease, and a signal would be improperly lost. Thus, the VLPA method requires that the 3′-most repeat of the target DNA anneals to the 5′-most repeat on the array probe (in a system where the probe is 5′ end attached to the array). To ensure that the target anneals to the probe in the proper register, a "clamp" sequence could be added to both the target and probe DNA. The clamp sequence is added at the microarray-proximal end of the probe, and its complement is added at the label-distal end of the target (Fig. 2B and D). The clamp sequence can be more GC-rich than the repeat sequences, thereby biasing the hybridization to the proper register (Fig. 2B and D). While it may be possible to distinguish between targets with different numbers of repeats without using the clamp, the addition of this clamp sequence to the method ensures that an obvious and measurable signal difference will be generated between positive and negative probes without having to resort to cumbersome and specialized hybridization conditions.

Demonstration of the Feasibility of VLPA

To demonstrate the feasibility of this new technique, we performed proof-of-principle experiments using commercially synthesized oligonucleotides of known identity and length (Table 1).
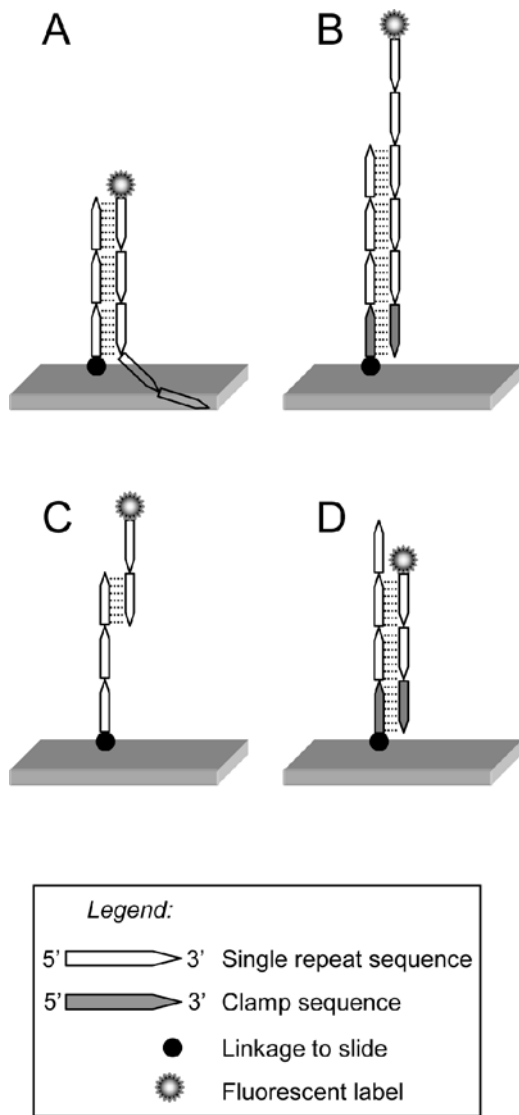
FIG. 2—*Use of the clamp sequence ensures proper hybridization of target to probe: (A) A target with more repeats than the probe attached to the microarray could anneal in an improper register such that fluorophore is not removed by nuclease. (B) Addition of a clamp sequence ensures that the target anneals in the proper register. (C) A target with fewer repeats than the probe anneals in an improper register and the fluorophore is improperly cleaved. (D) The clamp sequence ensures that the target anneals in the proper register.*

The target oligonucleotide was 5′ end-labeled with Cy5 and contained three repeats of 10 base pairs in length and a 3′ GC-rich 15 bp clamp sequence that was not complementary to the repeats. The probe oligonucleotides contained a 5′ amine group (to facilitate attachment to the microarray) followed by the complement of the clamp sequence and either one, two, three, four, or five tandem repeats. Arrays were printed with probe oligos and hybridized with target using conventional means. S1 endonuclease was used to remove ssDNA after hybridization. In each experiment, hybridizations were performed on three identical microarrays: one test array and two control arrays. The first control microarray served as a pre-nuclease incubation control and was processed and analyzed immediately after hybridization. The second control array was subjected to a post-hybridization incubation in S1 nuclease buffer without S1 nuclease and served as a control for the buffer and incubation conditions. The third test microarray was subjected to a post-hybridization incubation in S1 nuclease buffer containing S1 nuclease and was otherwise treated identically to the second microarray.

The fluorescence intensities for the control hybridization were similar between oligos with 1, 2, or 3 repeats (Table 2). Likewise, the fluorescence intensities of the features incubated in buffer without S1 nuclease were similar for 1, 2, or 3 repeats (Table 2). However, the fluorescent signal from the features with 1-repeat probes was substantially weaker than the signal from the features with 3-repeat probes on the microarray that was incubated in S1 nuclease. The features with 2-repeat probes showed a moderate decrease in signal relative to the 3-repeat probe. To quantitate the effects of the nuclease digestion on signals from the different probes, we analyzed four representative experiments that were performed identically but independently and calculated the mean fluorescence intensity from each probe as described in Materials and Methods.

On the two control arrays, the signal from the 1- and 2-repeat probes was not substantially reduced. In contrast, after S1 nuclease digestion, the signal from the 1-repeat probe was reduced approximately 5-fold compared to the signal from the 3-repeat probe ($p < 0.0001$), and the signal from the 2-repeat probe was reduced by about 38% ($p < 0.0001$) (Table 2). In other experiments, decreases in signal of as much as 20-fold have been observed from the 1-repeat probe (data not shown). No hybridization was observed of the target to a heterologous probe sequence (data not shown).

Analysis and Forensic Applications

In an effort to overcome the similar hybridization efficiency of STRs with similar lengths, we have designed VLPA, which incorporates a nuclease treatment and specialized clamp sequence to allow STR length determination. In our proof-of-principle experiments using a target sequence with three repeats, we observed a strong signal using the three-repeat probe, as expected. As predicted, the signal from the one-repeat probe was substantially reduced. The signal from the two-repeat probe was also reduced significantly, although more moderately. This may be due to non-linearity of fluorescent detection. Alternatively, the fluorophore could interfere with nuclease activity on the ssDNA in immediate proximity. The insertion of a spacer sequence between the repeats and the fluorophore of the target oligonucleotide might be a useful addition to enhance the robustness of the assay. Nonetheless, the observed decrease in the signal from the two-repeat probe was statistically significant.

These experiments demonstrate that the S1 nuclease treatment results in reduced signal from features with fewer repeats than the target. These data are consistent with the expected pattern of nuclease digestion and support the feasibility of the variable-length probe array STR profiling method. To our knowledge, this work represents the first selective digestion of single-stranded DNA on the surface of a microarray with an endonuclease.

The above experiments describe a process for determining the number of repeats of a single STR sequence. This method can be expanded to identify many different STR sequences on a single microarray in one experiment. Typical identification of a human being involves using 13 different STRs, each with 3-15 tandem repeats, in a profiling experiment. For each STR, all possible different lengths of probes must be represented as features on the microarray. Thus, as few as several hundred different features could be sufficient to uniquely identify an individual. Because current microarray technology allows hundred of thousands of unique features on a single chip, multiple copies of each feature can be incorporated into the

assay to ensure accuracy. Using a single microarray, thousands of identical features can be compared to each other to distinguish between datasets with slightly different average fluorescence levels. The microarray design can also incorporate a variety of controls of similar length and sequence to the relevant sequences to eliminate background signal and ensure accuracy in relating the fluorescence levels to repeat number.

A key innovation of the VLPA method is the use of the clamp sequence to prevent slippage and ensure proper hybridization. The sequences that flank the STRs in the human genome are the logical choice for these clamp sequences in practice.

In practice, several complicating issues may arise with forensic specimens. Many STR alleles contain a partial repeat or other variation of an adjacent set of exact tandem repeats. Other situations requiring special consideration are heterozygosity, mixtures, or any other case in which two or more target sequences are present in an unknown sample. In such cases, additional probe sequences would be added to the microarray to cover each example of a possible known variant, and cross hybridization issues would be avoided by using precise control of hybridization conditions. The addition of a microfluidics system to the VLPA method could allow us to vary experimental conditions such as temperature or buffer and to make comparisons between hybridizations under several different conditions within a single experiment.

Although the fluorescence detection system is sufficient for many purposes, applications requiring a high level of sensitivity and quantitation may benefit from the use of alternate technologies such as the Biomagnetic Gene Chip (MagArray^TM) developed by Stanford, which will ultimately allow detection of a single hybridization event on a microarray as well as accurate quantitation of the number of labels detected from a single feature over a range of about three orders of magnitude (11).

We anticipate that this technology could be easily adapted into portable, rapid detection systems for use in forensic identification and military applications in the field. Once integrated with a microfluidics system for sample preparation, hybridization, and enzymatic digestion, as well as an electronic system for detection readout, the entire system could ultimately be contained in a package the size of a laptop computer or handheld device.

The VLPA technology will be useful in a wide variety of applications that use STR analysis, including individual identification, paternity testing, and cancer diagnosis. We are currently developing conditions to adapt this technology for identification of human STRs.

References

1. The future of forensic DNA testing. Department of Justice, 2000, http://www.ncjrs.org/pdffiles1/nij/183697.pdf
2. Butler JM. Forensic DNA typing. London: Academic Press, 2001.
3. Gill P. Role of short tandem repeat DNA in forensic casework in the UK—past, present, and future perspectives. Biotechniques 2002; 32(2):366–85. [PubMed]
4. Hoyle R. The FBI's national DNA database. Nature Biotechnology 1998;16:987. [PubMed]
5. Goedecke N, McKenna B, El-Difrawy S, Carey L, Matsudaira P, Ehrlich D. A high-performance multilane microdevice system designed for the DNA forensics laboratory. Electrophoresis 2004;25(10–11):1678–86. [PubMed]
6. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B. J Validation of STR typing by capillary electrophoresis. Forensic Sci 2001;46(3):661–76.
7. Willse A, Straub TM, Wunschel SC, Small JA, Call DR, Daly DS, Chandler DP. Quantitative oligonucleotide microarray fingerprinting of *Salmonella enterica* isolates. Nucleic Acids Res 2004;32(5):1848–56. [PubMed]
8. Schena M. Microarray biochip technology. Natick, MA: Eaton Publishing, 200.
9. Stenirri S, Foglieni B, Manitto MP, Martina E, Brancato R, Cremonesi L, Ferrari M. Single nucleotide polymorphism and mutation identification by microelectronic chip technology. Minerva Biotecnologica 2002;14:241–6.
10. Radtkey R, Feng L, Muralhidar M, Duhon M, Canter D, DiPierro D, Fallon S, Tu E, McElfresh K, Nerenberg M, Sosnowski R. Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip. Nucleic Acids Res 2000;28(7):E17. [PubMed]
11. Li G, Joshi V, White RL, Wang SX, Kemp JT, Webb C, Davis RW, Sun S. Detection of single micron-sized magnetic bead and magnetic nanoparticles using spin valve sensors for biological applications. J Appl Phys 2003;93(10):7557–9.

Additional information and reprint requests:
Chris D. Webb, Ph.D.
Stanford University School of Medicine
300 Pasteur Drive
Alway Building, Room M-121, MC 5119
Stanford, CA 94305-5119
E-mail: cdwebb@stanford.edu