Chapter 10

DNA Microarray-Based Mutation Discovery and Genotyping

David Gresham

Abstract

DNA microarrays provide an efficient means of identifying single-nucleotide polymorphisms (SNPs) in DNA samples and characterizing their frequencies in individual and mixed samples. We have studied the parameters that determine the sensitivity of DNA probes to SNPs and found that the melting temperature (T_m) of the probe is the primary determinant of probe sensitivity. An isothermal-melting temperature DNA microarray design, in which the T_m of all probes is tightly distributed, can be implemented by varying the length of DNA probes within a single DNA microarray. I describe guidelines for designing isothermal-melting temperature DNA microarrays and protocols for labeling and hybridizing DNA samples to DNA microarrays for SNP discovery, genotyping, and quantitative determination of allele frequencies in mixed samples.

Key words: DNA microarray, Single-nucleotide polymorphisms, Isothermal melting temperature, SNPscanner, Bulk segregant mapping

1. Introduction

The original motivation for the development of DNA microarrays was the detection of genome sequence variation (1, 2). Although myriad applications of DNA microarrays were subsequently developed, including the analysis of mRNA expression levels (3), protein–DNA interactions (4), and genome amplifications and deletions (5, 6), the discovery and analysis of single-nucleotide polymorphism (SNP) variation using microarrays has remained a mainstay of modern molecular genetics.

All DNA microarray methods for detecting sequence differences rely on the chemistry of DNA duplex formation. Under appropriate reaction conditions, duplexes that are perfectly complementary in their DNA sequence are strongly favored over duplexes that contain one or more mismatched bases. In a typical

Virginie Orgogozo and Matthew V. Rockman (eds.), *Molecular Methods for Evolutionary Genetics*,
Methods in Molecular Biology, vol. 772, DOI 10.1007/978-1-61779-228-1_10,

experiment, the efficiency of duplex formation is measured by labeling a DNA sample with a fluorophore and quantifying the fluorescent signal at thousands to millions of probes following a hybridization reaction. Sample DNA fragments that are perfectly complementary to the probe sequence will exhibit maximal fluorescent signals, whereas the presence of even a single base difference that reduces complementarity results in diminished signals.

For the purpose of DNA sequence comparison, it is necessary to maximize the difference in efficiency of matched and mismatched duplex formation. Although several factors determine the efficiency of these reactions, two factors dominate. The first is the position of the mismatched base within the probe. Empirical studies have shown that mutations corresponding to mismatches in the central portion of the probe have the greatest impact on hybridization efficiency (7, 8). The second is the relationship between the predicted probe melting temperature (T_m) and the temperature at which the hybridization reaction is performed (9). The T_m of a probe is defined as the temperature at which 50% of the DNA molecules are in a duplex state. Previously, we performed a systematic study of the relationship between probe T_m and the discriminatory power of a probe sequence. We found that a probe melting temperature 2–5°C lower than the temperature at which hybridization is performed maximizes the sensitivity of duplex formation to single-base mismatches (9).

Here, I describe guidelines for designing DNA microarrays that employ these two simple principles. The manufacture of DNA microarrays corresponding to these designs is best realized using commercial manufacturers such as Agilent, Nimbelgen (Roche), or Affymetrix. These designs are well suited to SNP discovery on a genome scale (7) or targeted SNP genotyping of either individuals or pools of millions of individuals (10). I provide methodological details on preparation of samples and DNA hybridization experiments. The detailed steps are specific to the Agilent system, but should be readily adapted to other two-color hybridization platforms. In addition, I provide guidelines for analysis of hybridization data.

2. Materials

2.1. DNA Preparation and Labeling

1. High-quality genomic DNA is purified using Genomic Tip 100/G (Qiagen).
2. A sonicator is required for fragmenting DNA. Alternative means of random fragmentation are also possible, such as nebulization.
3. A Qubit Fluorometer (Invitrogen) provides the most accurate means of determining DNA concentration.

4. Quant-iT dsDNA Assay Kit, Broad Range (Invitrogen).
5. DNA Clean and Concentrator-5 columns (Zymo Research).
6. BioPrime Array CGH Genomic Labeling System (Invitrogen) contains all the components for labeling DNA except the cyanine-labeled nucleotides. These components include random primers (hexamers), a mix of nucleotides and the Klenow enzyme.
7. Cyanine-5 and cyanine-3 labeled dUTP (suppliers include Perkin Elmer and GE Healthcare).
8. 3 M sodium acetate, pH 5.2.
9. 100% ethanol.

2.2. DNA Hybridization and Washing

1. 2× Hi-RPM Hybridization Buffer, 25 ml (Agilent).
2. 10× GE Blocking Agent (lyophilized pellet) (Agilent).
3. Hybridization gasket slide (Agilent). Note that Agilent DNA microarrays come in a variety of formats ranging from 1 array/slide to 8 arrays/slide. It is important to purchase the appropriate corresponding hybridization gasket slide.
4. Express Plus 0.22-μm Stericup and Steritop filter (Millipore).
5. Wash A: 700 ml water, 300 ml 20× SSPE, and 0.25 ml 20% *N*-lauroylsarcosine. Filter Wash A using a Stericup 0.22-μm filter into a 1 liter bottle.
6. Wash B: 997 ml water, 3 ml 20× SSPE, and 0.25 ml 20% *N*-lauroylsarcosine. Filter Wash A using a Stericup 0.22-μm filter into a 1 liter bottle.
7. Acetonitrile. Acetonitrile is mildly toxic and should be handled with care.
8. Hybridization chambers and tweezers (Agilent) (see Note 1).
9. Four wash chambers and one rack. I use a 20-slide unit staining dish and slide rack from Electron Microscopy Sciences.
10. Four magnetic stir plates and mini stir bars.

3. Methods

3.1. Microarray Probe Design

Depending on whether one seeks to identify unknown SNPs by hybridizing to probes of known sequence or to genotype known SNPs in individual or mixed samples, two distinct microarray designs are required. In both cases, a reference genome sequence is required, and for genotyping microarrays the alternate allele for each SNP must be known. For SNP identification, probes are designed to overlap in a tiled manner. For genotyping, it is necessary to design probes in which the known SNP lies within the central region of

the probe. For both applications, an isothermal-melting probe design should be employed to maximize the potential for SNP identification or the discriminatory power of an allele-specific probe. I provide algorithmic guidelines, and examples in the notes, for simple computer code that must be written to generate either of these two microarray designs.

3.1.1. Isothermal Probe Design

1. The T_m (in °C) for a given DNA sequence is calculated using the relationship $T_m = \Delta H^\circ \times 1{,}000/(\Delta S^\circ + R \times \ln(C_T/x)) - 273.15$ where the $R = 1.9872$ cal/Kmol is the gas constant, $x = 4$ for nonself-complementary duplexes, and a total strand concentration (C_T) of 0.6×10^{-12} M is appropriate.
2. To calculate the T_m of a probe, it is necessary to determine the total enthalpy (ΔH°) and entropy (ΔS°) of the DNA sequence using the nearest neighbor parameters (Table 1). These parameters are determined for a salt concentration of 1 M NaCl, but they also provide suitable estimations for lower salt concentrations.
3. A target probe T_m should be selected. We have found that a T_m of 60°C results in probes of a median length of 25 bases and is a suitable target T_m for isothermal probe design.
4. Following completion of microarray probe design, each probe should be tested for uniqueness in the genome using a program such as BLAT. Nonunique probes should be excluded from the design. In addition, extremely short probes, which typically

Table 1
Nearest neighbor thermodynamic parameters for complementary dinucleotides in 1 M NaCl

Dinucleotide	ΔH° (kcal/mol)[1]	ΔH° (kcal/mol)[2]	ΔS° (e.u.[3])[1]	ΔS° (e.u.)[2]
AA	−8.4	−7.9	−23.6	−22.2
AT	−6.5	−7.2	−18.8	−20.4
TA	−6.3	−7.2	−18.5	−21.3
CA	−7.4	−8.5	−19.3	−22.7
GT	−8.6	−8.4	−23.0	−22.4
CT	−6.1	−7.8	−16.1	−21.0
GA	−7.7	−8.2	−20.3	−22.2
CG	−10.1	−10.6	−25.5	−27.2
GC	−11.1	−9.8	−28.4	−24.4
GG	−6.7	−8.0	−15.6	−19.9

[1]Values from (16)[1] and (15)[2] are provided. In our study of optimized microarray design, we used values from (16), but the values provided in (15) are improved estimations of the parameters. [3]Entropy units, which are equal to 1 Cal/K/mol

have an unusually high GC content, or extremely long probes, which have an unusually high AT content, are likely to have high signal variance in experiments. These probes typically have low sequence complexity and should be removed.

3.1.2. Probe Design for SNP Identification

1. Define the genomic region of interest. For microbial species with small genomes (<20 Mb), it is possible to design microarrays against the entire genome sequence (For example, with an average of 12-base spacing, a 20-Mb genome can be tiled with ~1.7 million overlapping probes). For larger genomes, a targeted approach is recommended, for example designing probes corresponding to the coding fraction of the genome.
2. Remove low complexity and repetitive DNA sequence elements from the genomic sequence. This includes telomeric DNA, retrotransposon sequence, and ribosomal DNA. SNP variation cannot be reliably detected in these regions using DNA microarrays.
3. Initiate the first probe sequence from 5′ end of selected genomic sequence. Continue incrementing probe sequence until the estimated melting temperature of the probe is minimally different from the target melting temperature (see an example of the probe selection process in Note 2).
4. Initiate the next tiled probe from the site corresponding to the center of the previous probe (see Note 3).
5. Continue probe design in this way until the region is completely covered with overlapping tiled probes of a uniform melting temperature.

3.1.3. Probe Design for SNP Genotyping

1. Identify genomic coordinates and the two alleles for each SNP. For biallelic polymorphisms, it is necessary to have two independent probes that differ at a single base complementary to the SNP. The local sequence flanking ~20 bases on each side of the probe must be known.
2. Initiate probe design from the SNP site incrementing the probe on alternating 5′ and 3′ sides of SNP until the estimated melting temperature of the probe is minimally different from the target melting temperature (see example in Note 4).
3. Repeat this procedure for the alternative allele for each SNP (as described in Note 4).

3.2. Hybridization Experiments

The hybridization method for performing SNP discovery or genotyping is based on a standard Agilent comparative genomic hybridization (CGH) protocol. High-quality DNA is purified and randomly fragmented using sonication. Labeling by incorporation of a cyanine-3 or cyanine-5 labeled nucleotide is performed using the Klenow DNA polymerase. The labeling reaction is performed

at a temperature that limits the processivity of the polymerase resulting in a uniform population of small fragment size. Short DNA fragments of <300 bp are required for efficient duplex formation. This labeling procedure results in an amplification of the sample DNA by approximately sevenfold. In my experience, this method does not result in biased amplification so long as the sample DNA has been randomly fragmented.

Two-color DNA microarray platforms allow for cohybridization of two samples enabling a variety of experimental designs. For mutation discovery, cohybridization of genomic DNA from a sample of known sequence (typically the sequence that was used for the microarray design) with a sample of unknown sequence is a common experimental design. For genotyping arrays two strains that are polymorphic at the targeted SNPs can be cohybridized. For bulk segregant mapping, one label should be used for the mixed sample and one label should be used for a heterozygous strain to enable comparison of the population allele frequency with a 50:50 allele frequency.

3.2.1. Preparation of Genomic DNA

1. Prepare genomic DNA using Qiagen Genomic Tip 100 following manufacturer's protocol (see Note 5).
2. Dilute 5 μg of sample genomic DNA in a total volume of 200 μl dH_2O (see Note 6).
3. Sonicate the DNA. We use the settings of power = 1, duration = 0.5 s, total = 15 s on a Misonix Sonicator 4000 Homogenizer.
4. Run 15 μl on a 1% agarose gel to confirm that the sonicated product is sufficiently fragmented. There should be a broad distribution of DNA fragments with a median size of ~600 bp.
5. Concentrate the DNA in a DNA Clean and Concentrator-5 column. Add 1 ml of DNA binding buffer to 200 μl of sonicated genomic DNA and load 600 μl onto column. Perform a quick hard spin (see Note 7), discard the flow-through, and repeat with the additional 600 μl.
6. Elute DNA from the column in 25 μl H_2O.
7. Determine the DNA concentration using a fluorometer (see Note 8). You should expect around 2 μg of DNA. You need at least 1 μg of fragmented DNA to proceed with labeling.

3.2.2. Labeling DNA with Cyanine-Labeled Nucleotides

1. In a 200 μl PCR tube bring 1,000 ng of DNA to 72 μl in H_2O. Place the tube in ice and add 60 μl 2.5× random primer solution, which is provided in the BioPrime Array CGH Genomic Labeling System.
2. Denature the DNA by heating for 8 min at 99°C in PCR block, then fast ramp cool down to 4°C for 8 min. Centrifuge briefly to recover all liquid.

3. Add 13 µl of 10× dUTP mixture (provided in the kit) to the tube followed by 2 µl Cy-labeled dUTP. Mix briefly. Add 3 µl of Klenow fragment and mix gently but thoroughly by pipetting.
4. Incubate at 25°C for 16 h in a PCR machine.
5. Add 15 µl of stop buffer (provided in the kit) to inactivate the reaction.
6. Transfer the solution to a 1.5-ml tube and add 16 µl of 3 M sodium acetate (pH 5.2). Mix the contents of the tube and then add 400 µl of ice-cold ethanol. Mix by inverting the tube and then place it at –20°C for 1–2 h.
7. Centrifuge at maximum speed for 10 min to pellet the precipitated DNA. Carefully remove the supernatant and then wash the pellet with 500 µl of 80% ethanol at room temperature. Centrifuge at maximum speed for 10 min and then carefully remove the supernatant.
8. Dry the pellet at room temperature by keeping the lid of the tube open (see Note 9).
9. Resuspend the pellet in 50 µl H_2O. Mix well by vortexing. It may be necessary to disturb the pellet with a pipette tip to ensure full resuspension.
10. Run 5 µl on a 2% agarose gel to confirm tightly distributed band around ~100 bp (Fig. 1). If the labeled sample does not resemble the example, repeat the labeling procedure.

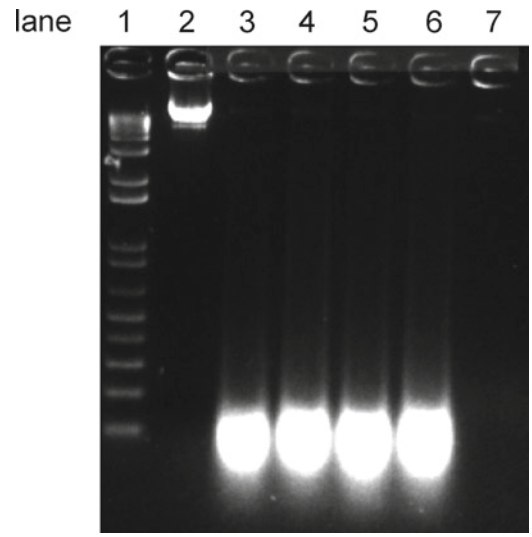

Fig. 1. Cyanine-dUTP labeling using Klenow enzyme at 25°C. Genomic DNA (lane 2) is fragmented using sonication to a median size of ~600 bp (not shown). The labeling procedure generates short fragments distributed around ~100 bp. Products from four independent labeling procedures are shown in lanes 3–6. Lane 1 contains a 1-kb ladder. Lane 7 is a blank control in which no template DNA was added to the labeling reaction.

11. Quantify the DNA using a fluorometer. You should expect a yield of approximately 5–7 μg.
12. The dye incorporation should be determined using a Nanodrop ND-1000, a spectrophotometer capable of quantifying cyanine-3 and cyanine-5 levels on the basis of absorbance. Our dye incorporation is typically greater than 10 pmol/μg DNA.
13. Store the labeled sample at –20°C.

3.2.3. Performing Microarray Hybridization

1. Add 1,350 μl of distilled water to the lyophilized pellet of 10× blocking agent. Leave for 60 min at room temperature to reconstitute. Store at –20°C thereafter. The solution can be reused.
2. For each hybridization reaction, add to a 1.5-ml tube 200 ng of each labeled DNA sample and water to a total volume of 208 μl (ensure that there are at least 2 pmol of dye in each channel). Then, add 52 μl of 10× Blocking Agent and 260 μl of 2× Hi-RPM Hybridization Buffer for a final volume of 520 μl. Mix sample well by vortexing. Heat at 95°C in a heat block for 5 min. Then, incubate at 37°C for 30 min. Perform a quick spin to collect the sample at bottom of tube.
3. Slowly dispense 490 μl of hybridization mixture onto a gasket slide (see Note 10).
4. Place a microarray slide on top of the gasket slide and create an enclosed reaction volume using the Agilent hybridization chambers.
5. Place the hybridization chamber in the hybridization oven at the appropriate hybridization temperature (e.g., hybridization should be performed at 62–65°C for a 60°C isothermal probe design) and a rotation speed of 20 rpm.
6. Allow hybridization to occur for 20 h.

3.2.4. Washing Microarray

1. Prepare four wash chambers, each containing a magnetic stirbar and place them on magnetic stirplates. Fill the first two chambers with wash A, the third chamber with wash B, and the fourth chamber with acetonitrile. Place the slide rack in the second chamber.
2. Disassemble each hybridization chamber one at a time. Place the microarray and gasket slide into the first chamber of Wash A. Use the plastic tweezers to gently wedge open the sandwich while keeping the microarray slide submerged in Wash A. You can allow the gasket slide to fall to the bottom of the wash chamber. Transfer microarray slide to the slide rack in the second Wash A chamber. Leave a gap between each slide and between the slides and the wall of the slide rack.

3. Once all the slides are in the rack, stir for 1 min in wash A. For all stirring steps, the wash liquid should be visibly turbulent so that a small vortex is visible. Make sure that the entire slide is submerged at all times.
4. Transfer the rack into Wash B and stir for exactly 1 min.
5. Transfer the rack into the acetonitrile. Leave submerged for 30 s. Slowly and evenly pull the rack out of the acetonitrile ensuring that all liquid drains from the slide
6. Dab microarray slide with a kimwipe to remove any remaining liquid.
7. Load the slides into scanning cartridges, with the Agilent name facing up and out. Do not touch anywhere except the edges of the slide and the bar code.

3.3. Data Analysis

3.3.1. Recommended Control Experiments

1. DNA-microarray-based SNP discovery is based on a relative decrease in hybridization efficiency with respect to that obtained for perfectly complementary DNA. Repeated hybridization experiments of a strain identical to the genome sequence used to design the microarray provide a means of establishing the expected signal at each probe for perfectly matched DNA.
2. Genotyping individuals on biallelic genotyping microarrays requires knowledge of the expected intensity at each allele-specific probe. Cohybridization of genomic DNA from two samples that are homozygous and carry the alternate allele at each locus provides the ideal means of determining these expected intensities.
3. Bulk segregant mapping requires estimation of allele frequencies in mixed populations on the basis of hybridization signal. By hybridizing a strain that is heterozygous at each locus, in addition to strains that are homozygous at each locus (in two different experiments), it is possible to generate a three-point calibration of the expected intensity at each probe for 0, 50, and 100% allele frequencies.

3.3.2. Detection of Point Mutations Using Tiled Overlapping Isothermal-Melting Probes

1. We have previously analyzed data for the purpose of mutation discovery using a likelihood ratio-based approach. This method requires a training dataset that consists of multiple hybridizations of the perfectly matched genome (to determine a mean and variance for each probe) and at least one hybridization experiment of a genome containing numerous known SNPs (several thousand) to train the algorithm. This method is analytically straightforward and readers should consult our previous publications for information on implementing this algorithm (7, 9).
2. If the aim is to identify SNPs for subsequent use as genetic markers, and, therefore, a level of false negatives can be toler-

ated, then a simple statistical test is likely to suffice for determining whether a hybridization signal from a sample of unknown genotype at a given probe is likely to be the result of imperfectly matched DNA sequence. One strategy entails comparing the hybridization signal at each probe to the triplicate hybridization data from the nonpolymorphic reference genome. This approach has been used for identifying single-feature polymorphisms (SFPs) (11, 12), and the reader should consult these references for details on this strategy.

3.3.3. Analysis of Genotyping Data

1. For each pair of probes interrogating a biallelic SNP, the aim is to assess whether an individual is homozygous for either allele or heterozygous at the locus. Several statistical procedures have been developed for this purpose including likelihood-based (13) and classification-based methods (14).
2. For bulk segregant mapping, the relative intensity at the two allele-specific probes for each locus can be converted to an estimated allele frequency. The reader should consult (10) for details on this method of anaylsis.

4. Notes

1. Agilent microarrays are designed to work with Agilent hybridization chambers and oven. I have not tested whether other hybridization chambers and ovens are compatible with Agilent microarrays.
2. An example of the method for identifying a tiling DNA microarray probe, for the purpose of SNP discovery, with a T_m nearest to 60°C given the target sequence: ATGGATTCTGAGGTTGCTGCTTTGGTTATTGATAACG. The T_m for each candidate probe is calculated by summing all the nearest neighbor values to calculate the total enthalpy and entropy of the DNA sequence using the values from (15). This calculation is performed iteratively for each candidate probe until the probe that is closest to the desired T_m is identified. In this particular example, the candidate probe of 23 nucleotides long should be selected because the T_m of this probe (59.4°C) is closer to the target T_m than the 24-nucleotide candidate probe (61.08°C).

Candidate probe sequence	Probe length (nucleotides)	Total $\Delta H°$(kcal/mol)	Total $\Delta S°$ (e.u.)	T_m (°C)
ATGGATTCTGAGGTTG	16	−120.3	−324.1	41.13
ATGGATTCTGAGGTTGC	17	−130.1	−348.5	46.37
ATGGATTCTGAGGTTGCT	18	−137.9	−369.5	48.91
ATGGATTCTGAGGTTGCTG	19	−146.4	−392.2	51.55
ATGGATTCTGAGGTTGCTGC	20	−156.2	−416.6	55.50
ATGGATTCTGAGGTTGCTGCT	21	−164	−437.6	57.31
ATGGATTCTGAGGTTGCTGCTT	22	−171.9	−459.8	58.40
ATGGATTCTGAGGTTGCTGCTTT	23	−179.8	−482	59.40
ATGGATTCTGAGGTTGCTGCTTTG	24	−188.3	−504.7	61.08

3. We have found that to maximize the sensitivity of overlapping tiling arrays for detection of SNPs it is necessary that every nucleotide fall within the internal 70th percentile of at least one probe. To achieve this a good rule of thumb is to start the subsequent probe from the center of the previous probe. In the example provided in Note 2, the subsequent probe should be initiated at the twelfth nucleotide of the 23-mer probe that was identified as having a T_m closest to the target T_m of 60°C. By following the same design guidelines the subsequent probe is the 25-mer GGTTGCTGCTTTGGTTATTGATAAC, which has a T_m of 59.23°C. The next probe should now commence at the 13th base of this probe and so on.
4. An example of the method for identifying two probes with melting temperatures nearest to 60°C, when the SNP and adjacent sequence are known, is shown below. The example sequence is: ATGGATTCTGAGGTTGC[G/T]GCTTTGG-TTATTGATAACG. The nearest neighbor parameters from (15) were used.

Allele	Candidate probe sequence	Probe length (nucleotides)	Total $\Delta H°$(kcal/mol)	Total $\Delta S°$ (e.u.)	T_m (°C)
G	AGGTTGC*G*GCTTTGGT	16	−127.3	−334.5	50.62
	GAGGTTGC*G*GCTTTGGT	17	−135.5	−356.7	53.06
	GAGGTTGC*G*GCTTTGGTT	18	−143.4	−378.9	54.56
	TGAGGTTGC*G*GCTTTGGTT	19	−151.9	−401.6	56.87
	TGAGGTTGC*G*GCTTTGGTTA	20	−159.1	−422.9	57.22
	CTGAGGTTGC*G*GCTTTGGTTA	21	−166.9	−443.9	58.94
	CTGAGGTTGC*G*GCTTTGGTTAT	22	−174.1	−464.3	59.75
	TCTGAGGTTGC*G*GCTTTGGTTAT	23	−182.3	−486.5	61.24
T	AGGTTGC*T*GCTTTGGT	16	−125	−331.1	47.54
	GAGGTTGC*T*GCTTTGGT	17	−133.2	−353.3	50.17
	GAGGTTGC*T*GCTTTGGTT	18	−141.1	−375.5	51.83
	TGAGGTTGC*T*GCTTTGGTT	19	−149.6	−398.2	54.29
	TGAGGTTGC*T*GCTTTGGTTA	20	−156.8	−419.5	54.76
	CTGAGGTTGC*T*GCTTTGGTTA	21	−164.6	−440.5	56.59
	CTGAGGTTGC*T*GCTTTGGTTAT	22	−171	−460.9	57.9
	TCTGAGGTTGC*T*GCTTTGGTTAT	23	−180	−483.1	59.09
	TCTGAGGTTGC*T*GCTTTGGTTATT	24	−187.9	−505.3	60.02
	TTCTGAGGTTGC*T*GCTTTGGTTATT	25	−195.8	−527.5	60.88

In the case of the G allele-specific probe, the 22-mer probe (CTGAGGTTGC*G*GCTTTGGTTAT) should be chosen. For the T allele-specific probe, the 24-mer probe (TCTGAGGTTGC*T*GCTTTGGTTATT) should be selected.

5. The DNA columns invariably become clogged. Use of a modified rubber stopper with a single inlet to apply gentle air pressure (from house air or a syringe) expedites the flow.
6. All water used in this protocol is distilled and autoclaved.
7. All spins in this protocol are performed in a microcentrifuge at the maximum speed (e.g., 14,000 × *g*).
8. Measuring DNA concentrations using a fluorometer is much more accurate than on a spectrophotometer, which cannot distinguish DNA and contaminating RNA. A fluorometer should always be used in this protocol.
9. The dye cyanine-5 is particularly susceptible to degradation by ozone at levels as low as 5–10 parts per billion. This problem varies between labs and locations. The problem is particularly acute when the dried slide is exposed to ozone after the washes. A commercially available plastic barrier (Ozone Barrier Slide Cover; Agilent) can be placed over the slide once it is in the scanner cartridge to provide some protection from atmospheric ozone. An ozone-free workspace for washing and handling microarrays and labeled samples can be established using a NoZone WS Workpace (SciGene), which consists of a benchtop polycarbonate enclosure with an external ozone filtration system. If rapid degradation of cyanine-5 continues to be a problem,

potential solutions include the installation of carbon filters into the laboratory air handling system or the use of alternative dyes.

10. This protocol describes a hybridization reaction for a single microarray per glass slide. If other formats are used such as the 4×44 k or 8×15 k platforms, the total reaction volume should be scaled according to the reaction volume guidelines provided by Agilent. The amount of DNA should also be reduced by the appropriate factor.

Acknowledgments

I thank the labs of David Botstein, Leonid Kruglyak, Maitreya Dunham, and Justin Borevitz where many of these methods were developed. I also thank Bo Curry, Leonardo Brizuela, and Ben Gordon at Agilent Technologies for participation in the initial study of microarray design.

References

1. Maskos U, Southern, EM (1993) A novel method for the parallel analysis of multiple mutations in multiple samples. Nucleic Acids Res **21**:2269–2270
2. Gresham D, Dunham MJ, Botstein D (2008) Comparing whole genomes using DNA microarrays. Nat Rev Genet 9:291–302
3. DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278:**680–686
4. Lieb JD, Liu X, Botstein D et al (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet **28**:327–334
5. Pollack JR, Perou CM, Alizadeh AA, et al (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet **23**:41–46
6. Pinkel D, Segraves R, Sudar D, et al (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet **20**:207–211
7. Gresham D, Ruderfer DM, Pratt SC, et al (2006) Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. Science **311**:1932–1936
8. Ronald J, Akey JM, Whittle J, et al (2005) Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays. Genome Res **15**:284–291
9. Gresham D, Curry B, Ward A, et al (2010) Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. Proc Natl Acad Sci USA **107**:1482–1487
10. Ehrenreich IM, Torabi N, Jia Y, et al (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature **464**:1039–1042
11. Winzeler EA, Richards DR, Conway AR, et al (1998) Direct allelic variation scanning of the yeast genome. Science **281**:1194–1197.
12. Borevitz JO, Liang D, Plouffe D, et al (2003) Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res **13**:513–523
13. Cutler DJ, Zwick ME, Carrasquillo MM, et al (2001) High-throughput variation detection and genotyping using microarrays. Genome Research **11**:1913–1925
14. Liu WM, Di X, Yang G, et al (2003) Algorithms for large-scale genotyping microarrays, Bioinformatics **19:**2397–2403
15. SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA **95**:1460–1465
16. SantaLucia J, Allawi HT, Seneviratne PA (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry **35**:3555–3562