

- Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., and Sellers, W. R. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**, 203–209.
- Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel, D., and Albertson, D. G. (2001). Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* **29**, 263–264.
- van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009.
- Van Gelder, R. N., von Zastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D., and Eberwine, J. H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* **87**, 1663–1667.
- van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Bernards, R., and Friend, S. H. (2003). Expression profiling predicts outcome in breast cancer. *Breast Cancer Res.* **5**, 57–58.
- van't Veer, L. J., Paik, S., and Hayes, D. F. (2005). Gene expression profiling of breast cancer: A new tumor marker. *J. Clin. Oncol.* **23**, 1631–1635.
- Wang, D., Coscoy, L., Zylberberg, M., Avila, P. C., Boushey, H. A., Ganem, D., and DeRisi, J. L. (2002). Microarray-based detection and genotyping of viral pathogens. *Proc. Natl. Acad. Sci. USA* **99**, 15687–15692.
- Waring, J. F., Ciurlionis, R., Jolly, R. A., Heindel, M., and Ulrich, R. G. (2001a). Microarray analysis of hepatotoxins *in vitro* reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol. Lett.* **120**, 359–368.
- Waring, J. F., Jolly, R. A., Ciurlionis, R., Lum, P. Y., Praestgaard, J. T., Morfitt, D. C., Buratto, B., Roberts, C., Schadt, E., and Ulrich, R. G. (2001b). Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharmacol.* **175**, 28–42.
- Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O., and Schoolnik, G. K. (1999). Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc. Natl. Acad. Sci. USA* **96**, 12833–12838.

[3] Illumina Universal Bead Arrays

By JIAN-BING FAN, KEVIN L. GUNDERSON, MARINA BIBIKOVA,
JOANNE M. YEAKLEY, JING CHEN, ELIZA WICKHAM GARCIA,
LORI L. LEBRUSKA, MARC LAURENT, RICHARD SHEN, and DAVID BARKER

Abstract

This chapter describes an accurate, scalable, and flexible microarray technology. It includes a miniaturized array platform where each individual feature is quality controlled and a versatile assay that can be adapted for various genetic analyses, such as single nucleotide polymorphism genotyping,

DNA methylation detection, and gene expression profiling. This chapter describes the concept of the BeadArray technology, two different Array of Arrays formats, the assay scheme and protocol, the performance of the system, and its use in large-scale genetic, epigenetic, and expression studies.

Introduction

DNA polymorphisms and differences in gene expression provide the genetic basis for phenotypic variation. The systematic screening of an organism's genome for genetic variants and identification of environmental conditions that affect gene regulation and linking the former to a possible phenotype may advance our understanding of the underlying genetic or epigenetic mechanisms and lead to the identification of alleles (or haplotypes) and genes that contribute to phenotypic diversity. However, any comprehensive genome-wide association studies would need to genotype a large number of single nucleotide polymorphisms (SNPs) (or measure expression of a large number of genes) in large sample sets (Botstein and Risch, 2003), thus requiring a system that combines very high throughput and accuracy with very low cost per SNP (or gene) analysis (Kwok and Chen, 2003; Syvanen, 2005). The technology described in this chapter provides a solution to these challenges by integrating a miniaturized array platform with a high level of assay multiplexing and scalable automation. The system uses the Sentrix BeadArray technology in combination with a solid-phase, allele-specific extension, followed by polymerase chain reaction (PCR) amplification to achieve high multiplex levels. The system is able to maintain high assay multiplex levels and high sample throughput in an informatically integrated production environment. Gunderson *et al.* (2005) described a genome-wide scalable genotyping assay, the Infinium assay, which is also deployed on Sentrix arrays and allows simultaneous whole genome genotyping of over 100,000 SNP markers.

Material and Methods

BeadArray Technology

Conventional microarrays are manufactured by spotting or synthesizing probes onto two-dimensional substrates at known locations (Fodor *et al.*, 1991; Holloway *et al.*, 2002; Schena *et al.*, 1995). In contrast, BeadArray technology is based on the random self-assembly of a bead pool onto a patterned substrate (Fan *et al.*, 2003; Michael *et al.*, 1998; Oliphant *et al.*, 2002). As a part of the array manufacturing process, a decoding process is implemented to map the precise location of a specific bead type on

the array (Gunderson *et al.*, 2004). The number of independent array features is dependent on the complexity of the assay pool. A high degree of miniaturization can be achieved by exploiting the intrinsic size of the beads and patterned substrate. For example, the density of a randomly assembled 300-nm-diameter bead array is $\sim 40,000$ times higher than a typical spotted microarray (Michael *et al.*, 1998).

In the past, bead assemblies were encoded in a number of ways. The simplest method involved impregnation or attachment of dyes to the bead. Initial attempts to determine the location and identity of beads using this approach were largely affected by the variability in quantitation, instability, and array-to-array variation of the dye, among other technical problems, which limited the usefulness of the approach (Michael *et al.*, 1998). A number of other schemes that aim to encode particles directly with combinatorial codes generated by mixtures or spatial arrangements of optical signaling molecules have similar issues (Braeckmans *et al.*, 2002; Chan *et al.*, 2002; Fulton *et al.*, 1997; Han *et al.*, 2001; Lockhart and Trulson, 2001; Nicewarner-Pena *et al.*, 2001).

We sidestepped the need for complex dye chemistries and painstaking labeling processes by devising a novel, highly efficient decoding, rather than encoding, strategy that employs the specificity and reversibility of DNA hybridization (Gunderson *et al.*, 2004). Quantitatively pooled bead libraries are randomly self-assembled into etched microwell substrates (Barker *et al.*, 2003) (Fig. 1). The physical location or “map” of the beads is determined by serial hybridization steps using fluorescently labeled complementary oligonucleotides. The advantage of this approach is that dense packing with unlimited scalability can be achieved concomitantly with increased reliability. Specifically, each bead is derivatized on its surface with approximately 700,000 covalently attached oligonucleotide probes of a unique sequence (Kuhn *et al.*, 2004). Dye-labeled oligonucleotides then hybridize to the address sequence of each probe type in the hybridization or “decoding” process. The physical location of each of the bead types is thus mapped by using a decoding algorithm (Gunderson *et al.*, 2004) that tracks oligonucleotides successfully hybridizing to a specific address. With this compiled information, the physical location of each bead is mapped. The decoding process also serves as a quality control step for each feature of each array.

The address sequences are designed computationally to have no significant homology to genomic sequences within the species queried. With these “universal arrays,” the address sequences, called IllumiCodes, are used for decoding as well as analytical readout in genotyping or gene expression applications. Currently, up to 1624 unique bead types are represented in each of the universal arrays, with an average 30-fold bead type redundancy in each array ($\sim 50,000$ beads).

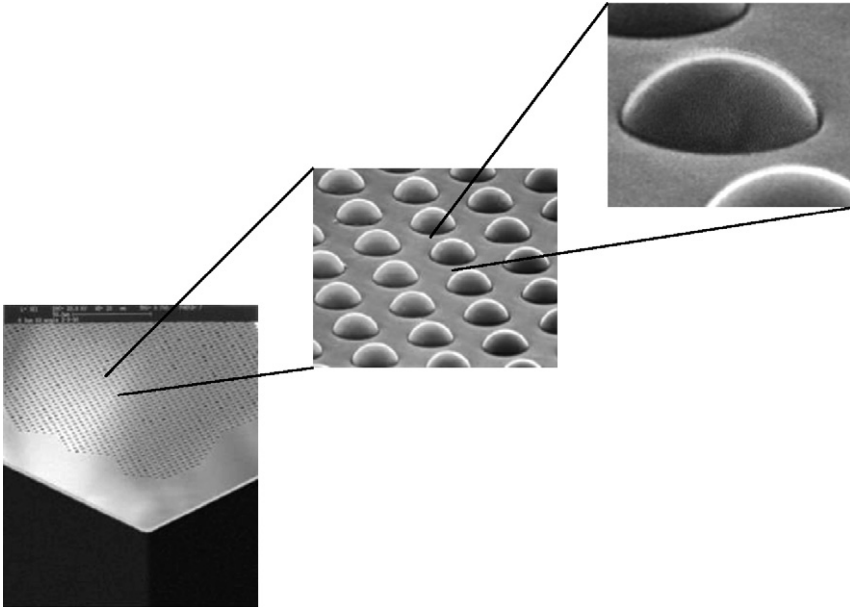


FIG. 1. BeadArray technology. Silica beads, each $3\ \mu\text{m}$ in diameter, are assembled randomly into wells with $5\text{-}\mu\text{m}$ center-to-center spacing. Each bead is represented with an average 30-fold redundancy on each array.

Array of Arrays Format

Illumina has developed two different Array of Arrays formats: the Sentrix Array Matrix (SAM) and the Sentrix BeadChip (Fig. 2). The SAM is an 8×12 matrix consisting of 96 fiber bundles (i.e., 96 independent arrays) with $\sim 50,000$ light-conducting strands per bundle. Thus, one SAM can assay 96 separate samples simultaneously. The SAM is compatible with standard microtiter plates used to carry out sample preparation and enzymatic steps prior to exposure of the beads in the array to the sample. Therefore, the platform can be readily incorporated into automated routines using standard robotic equipment.

The second array format, termed the Sentrix BeadChip, is a silicon substrate with the dimensions 2.5×8.25 cm. This platform is highly flexible and will allow configurations that assay 1 to 16 samples at a time. For example, our Infinium Whole Genome Genotyping Assay (100 K genotypes) is performed on a single BeadChip. If the BeadChip is formatted differently, such as in the Human-6 BeadChip for whole genome gene expression analysis, 6 samples may be analyzed on one BeadChip. The



FIG. 2. Two different Array of Arrays formats: the 96-sample Sentrix Array Matrix (SAM) and the multisample Sentrix BeadChip. The SAM substrate (upper) is composed of 96 arrays. Each array is composed of ~50,000 fiber-optic strands chemically etched to create a well for a single bead. The BeadChip formats (lower) feature flexible configuration with dense packing ratios that can be used for 1 to 16 samples per BeadChip.

16-sample BeadChip has the same feature-to-feature spacing as standard multichannel pipettes.

GoldenGate Genotyping Assay

The GoldenGate genotyping assay ([Fan *et al.*, 2003](#); [Shen *et al.*, 2005](#)) provides a flexible, accurate, and high-throughput SNP genotyping system for large-scale genetic analysis ([Fig. 3](#)). It allows for a high degree of locus multiplexing through highly specific extension and amplification steps. The multiplexed assay measures from 384 to 1536 SNPs in a single reaction. For example, a single SAM can be used to genotype 1536 SNPs across 96 samples, allowing the researcher to determine up to 150,000 genotypes simultaneously. Proprietary OligoDesigner software has been used to design hundreds of thousands of GoldenGate genotyping assays. It evaluates sequences flanking the targeted SNP, such as degree of homology to repeated sequences across the genome, palindromic sequences, GC&AT

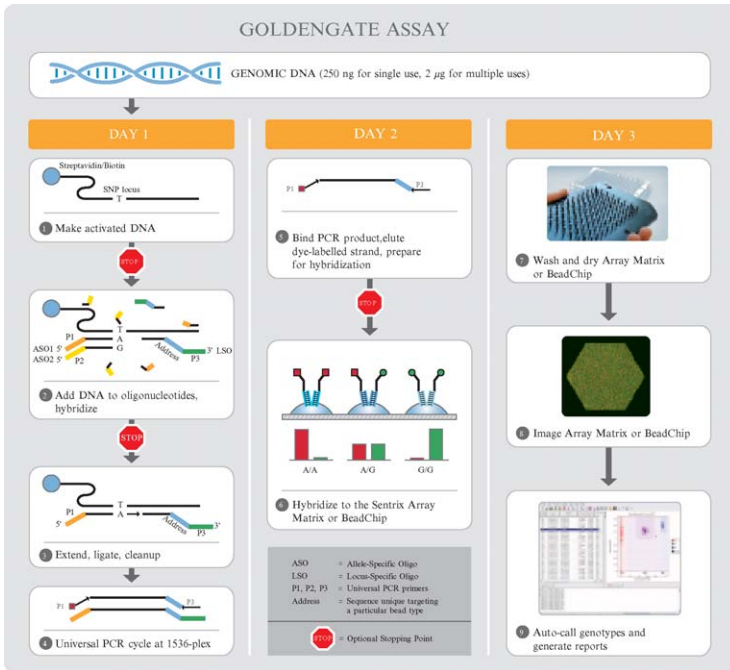


FIG. 3. Illustration of the GoldenGate genotyping assay process.

content, and neighboring polymorphisms. The allele-specific sequence is designed at a T_m of 60° ($57-62^\circ$), whereas the locus-specific probe is designed at a T_m of 57° ($54-60^\circ$). To assure high-quality data, various internal assay controls are used to assess the GoldenGate assay and array hybridization at various experimental steps, including gDNA/oligonucleotide annealing, PCR, array hybridization, and imaging. For example, controls include G/T or G/G mismatches; amplification balance of the two types of universal PCR primers; hybridization controls; and a double label control to estimate the optical balance of Cy3/Cy5 channels.

This section describes the standard GoldenGate genotyping assay procedures, although details of the protocol may vary depending on the specifics of the genotyping system used. All robotic processes were performed on a Tecan Genesis Workstation 150 (Tecan).

Biotinylation of Genomic DNA (Step 1). Genomic DNA ($5\ \mu\text{l}$ at $50\ \text{ng}/\mu\text{l}$) is mixed with $5\ \mu\text{l}$ biotinylation reagent (MS1; Illumina) and incubated at 95° for 30 min. A precipitation reagent (PS1; Illumina; $5\ \mu\text{l}$) is then added, followed by the addition of $15\ \mu\text{l}$ of 2-propanol. The samples are

mixed thoroughly and spun at high speed (3000g). The pellets are resuspended in buffer (RS1; Illumina; 10 μ l) and used in the subsequent assay step.

Annealing of Assay Oligonucleotides to Genomic DNA (Step 2). Genotyping assay oligonucleotides (corresponding to 384–1536 specific SNPs) and paramagnetic particles are then combined with the activated DNA in the oligonucleotide/target annealing step, in which the query oligonucleotides hybridize to the genomic DNA as it binds to paramagnetic particles. Two allele-specific oligonucleotides (ASOs) and one locus-specific oligonucleotide (LSO) are designed for each SNP (Fig. 3) (Fan *et al.*, 2003). All three oligonucleotide sequences contain regions of genomic complementarity and universal PCR primer sites; the LSO also contains a unique IllumiCode sequence complementary to a particular bead type. The IllumiCode sequence hybridizes to the universal bead type probes in the last hybridization step of the genotyping assay (see Step 6). The IllumiCodes represented on the universal array comprises 1624 artificial sequences carefully selected to not cross-hybridize with each other or with sequences in the human genome. LSOs are synthesized with a 5' phosphate to enable ligation. Annealing reagent (OB1; Illumina; 30 μ l) and SNP-specific oligonucleotides (OPA; Illumina; 10 μ l) are combined with DNA (10 μ l) to a final volume of 50 μ l. Annealing is carried out by ramping the temperature from 70 to 30° over ~2 h and then holding at 30° until the next processing step. Up to 1536 SNPs may be assayed simultaneously in this manner.

Assay Oligonucleotides Extension and Ligation (Step 3). Following assay oligonucleotide hybridization, several wash steps are performed to remove excess and mishybridized oligonucleotides. In a single reaction, a DNA polymerase with high specificity for a 3' match and no strand displacement or exonuclease activity is used to extend the ASO(s) that perfectly matches the target sequence at the SNP site and fill the gap between the ASO and the LSO (Step 3), and a DNA ligase is used to seal the nick between the extended ASO and the LSO to form PCR templates that can be amplified with universal PCR primers (Step 4). A master mix for extension and ligation (MEL; Illumina; 37 μ l) is added to the beads. Extension and ligation are carried out at 45° for 15 min. High locus specificity is achieved by the requirement that both the ASO and the LSO oligonucleotides hybridize to the same target site; extension of the appropriate ASO and ligation of the extended product to the adjacent LSO join information about the genotype present at the SNP site to the address sequence on LSO.

PCR Amplification (Step 4). After extension and ligation, the beads are washed with universal buffer 1 (UB1; Illumina) to reduce noise by removing excess and mishybridized oligonucleotides. The beads are then resuspended in 35 μ l elution buffer (IP1; Illumina) and heated at 95° for 1 min to

release the ligated products. The supernatant is then used in a 60- μ l PCR reaction. PCR reactions are thermocycled as follows: 10 min at 37°; 34 cycles (35 s at 95°, 35 s at 56°, 2 min at 72°); 10 min at 72°; and cooled to 4° for 5 min. The three universal PCR primers (P1, P2, and P3) are 5' labeled with Cy3, Cy5, and biotin, respectively. If necessary, uracil DNA glycosylase (UDG) can be added to the PCR master mix to help prevent PCR product contamination.

PCR Product Preparation (Step 5). Double-stranded PCR products are immobilized onto paramagnetic particles by adding 20 μ l of paramagnetic particle b reagent (MPB; Illumina) to each 60- μ l PCR reaction and then incubated at room temperature for a minimum of 60 min. Bound PCR products are washed with universal buffer 2 (UB2; Illumina) and denatured by adding 30 μ l 0.1 N NaOH. After 1 min at room temperature, the released single-stranded (ss) DNAs are neutralized with 30 μ l of hybridization reagent (MH1; Illumina) and hybridized to the universal array.

Array Hybridization (Step 6), Wash (Step 7), and Imaging (Step 8). The single-stranded, dye-labeled DNA products are hybridized to their complement bead type through their unique IllumiCode sequences (Step 6). The SAM or BeadChip is then washed (Step 7) and imaged (Step 8). Specifically, arrays are hydrated in UB2 for 3 min at room temperature and then preconditioned in 0.1 N NaOH for 30 s. Arrays are returned to the UB2 reagent for at least 1 min to neutralize the NaOH. The pretreated arrays are exposed to the labeled ssDNA samples described earlier. Hybridization is conducted under a temperature gradient program from 60 to 45° over ~12 h. The hybridization is held at 45° until the array is processed (Step 6). After hybridization, the arrays are first rinsed twice in UB2 and once with IS1 (IS1; Illumina) at room temperature with mild agitation (Step 7), dried for 20 min, and then imaged at a resolution of 0.8 μ m using a BeadArray reader (Illumina; [Barker et al., 2003](#)) (Step 8). PMT settings are optimized for dynamic range, channel balance, and signal-to-noise ratio. Cy3 and Cy5 dyes are excited by lasers emitting at 550 and 630 nm, respectively. A scan of the 96 arrays in a SAM takes approximately 90 min.

Automatic Genotype Scoring (Step 9). Based on the intensities detected from the two channels for the two respective alleles of each SNP, genotypes are called automatically using the proprietary genotyping software of Illumina (GenCall). GenCall software considers multiple factors; one of the first is the distribution of beads of the same type. Outliers are rejected to ensure genotyping accuracy.

A “quality” score, the GenCall score, is calculated for each genotype call, reflecting the degree of separation between homozygote and

heterozygote clusters for that SNP and the placement of the individual call within a cluster. It ranges from 0 to 1 and has been shown to correlate with the overall accuracy of the genotyping call (Fan *et al.*, 2003). Figure 4 shows four examples of GenCall software-produced data plots. These are genotype clusters generated from 90 individual DNA samples from a set of families affected by bipolar disorder, assayed with an oligonucleotide pool corresponding to ~1400 SNPs. Separation between homozygote and heterozygote clusters is excellent and variation within each cluster is small.

Results and Discussion

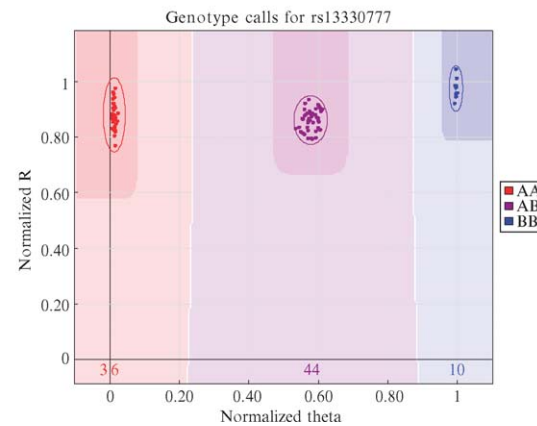
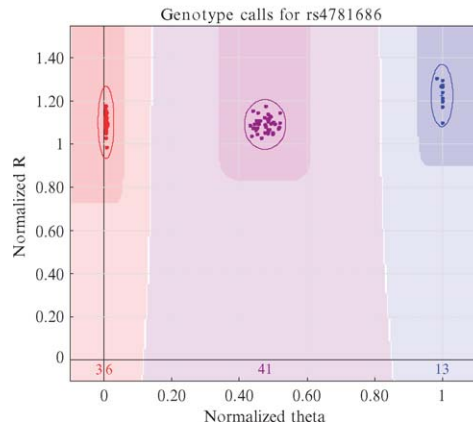
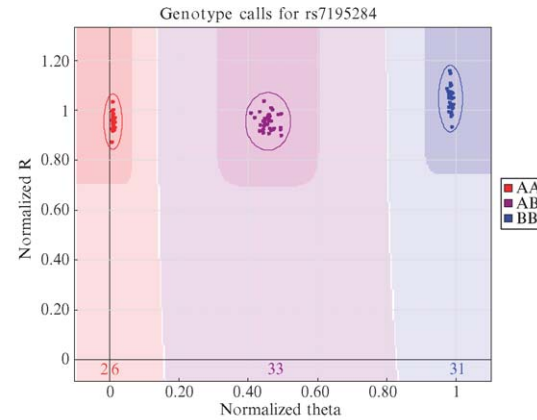
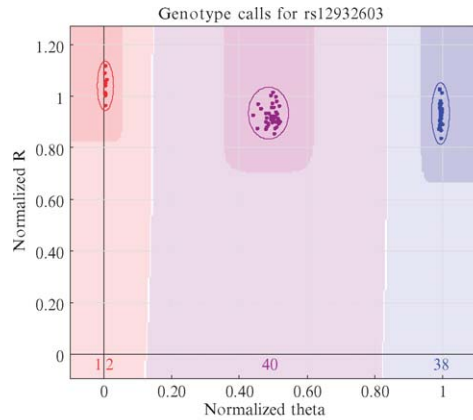
Assay Flexibility

One key aspect of the GoldenGate genotyping assay design is incorporation of a universal IllumiCode sequence, allowing the assay products to be read out on a universal array (Chen *et al.*, 2000; Fan *et al.*, 2000; Gerry *et al.*, 1999; Iannone *et al.*, 2000). This approach offers substantial flexibility for assay content development. Any custom set of assays can be made on demand simply by building the complement of the IllumiCode into the SNP-specific assay oligonucleotides (see Fig. 3, step 2). In addition, use of a universal array greatly simplifies the manufacturing process and reduces the assay development time (because there is no need to create custom arrays) and overall genotyping costs.

Genotyping Throughput

The ability to use either the Sentrix BeadChip or the SAM (Fig. 2) allows the user to perform as few as 6000 genotype calls (384 loci multiplex on a 16-sample BeadChip) to as many as 300,000 genotype calls (1536 loci multiplex on two 96-sample array matrices) in a single day. Automating the GoldenGate assay with a few liquid handling robots and a laboratory information management system allows throughput of over 1.6 million genotype calls per day.

An international consortium of leading investigators (including Illumina scientists) was formed to define a haplotype map for the human genome at a 5-kb resolution. The International Haplotype Map (HapMap) project aims to genotype more than 1 million SNPs in 288 individuals (Caucasians, Africans, and Asians) and to identify “tag” SNPs that represent the common haplotypes in the general populations (The International HapMap Consortium, 2003). The GoldenGate genotyping platform of Illumina is used by the consortium to genotype ~65% of the SNP loci in this project. This effort, along with others (Hinds *et al.*, 2005), provides the tools and



resources required for understanding the causal role of common human DNA variations in complex human traits.

Genotyping Performance

To date, over 500 million high-quality genotypes have been generated using the GoldenGate assay. About 1 million SNP loci have been designed and 88% of the attempted assays were successful. The assay development success rate is highly dependent on the SNPs selected. Of the DNAs processed, 97.8% were genotyped successfully and high-quality data were obtained. A primary reason for the failure to generate genotype calls successfully was insufficient DNA quantity. The genotyping result can be compromised significantly when less than 20% of the required input DNA (250 ng) is used. In general, the GoldenGate genotyping assay delivers a high call rate (99.8%) and call accuracy (>99.9%).

In addition, the GoldenGate assay can tolerate a certain degree of DNA degradation. We have obtained highly accurate genotyping results with whole genome-amplified DNA samples (Barker *et al.*, 2004) and DNA samples isolated from formalin-fixed, paraffin-embedded (FFPE) tissue samples (Fan *et al.*, unpublished data).

Implementation of the Methylation Profiling Assay on the SNP Genotyping Platform

We adapted the GoldenGate SNP genotyping system for high-throughput DNA methylation detection based on genotyping of bisulfite-converted genomic DNA. In this assay, nonmethylated cytosines (C) are converted to uracil (U) when treated with bisulfite, while methylated cytosines remain unchanged. The hybridization behavior of uracil is similar to that of thymine (T). Detection of the methylation status of a particular cytosine is carried out using a genotyping assay for a C/T polymorphism. The methylation status of the interrogated locus can be determined by

FIG. 4. GenCall software-produced plots of four randomly selected loci of 90 samples in polar coordinate representation. Each image is a graph of a single locus (assayed simultaneously with ~1400 other loci) with 90 “dots” representing individual DNA samples. The y axis is normalized intensity (sum of intensities of the two channels), and the x axis is the “theta” value ($\frac{2}{\pi} \tan^{-1} (Cy5/Cy3)$). Theta values near 0 (left side of graph) are homozygotes for allele “A,” and theta values near 1 (right side of graph) are homozygotes for allele “B.” The GenCall software automatically grouped the 90 DNAs for each locus into two homozygote clusters (left and right) and the heterozygote cluster (middle).

calculation of the ratio of the fluorescent signals from the “C” (methylated) and “T” (unmethylated) alleles.

We have characterized this assay system extensively. We established a standard process for probe design and data analysis, a standard bisulfite conversion protocol, and a set of internal controls and reference samples for assay development and calibration. We demonstrated that the assay is sufficient to detect changes in methylation status at 1536 different CpG sites simultaneously using 200 ng of human genomic DNA; it can detect as little as 2.5% of methylated DNA target in a complex sample and can distinguish a 17% methylation difference between samples with 95% confidence (Bibikova *et al.*, unpublished data). We also observed high concordance between results obtained by our microarray-based methylation analysis and those by traditional methods such as methylation-specific PCR and bisulfite sequencing. This technology should prove useful for comprehensive methylation analyses in large populations.

High-Throughput Gene Expression Profiling on Universal Array Matrices

We have developed a flexible, sensitive, accurate, and cost-effective gene expression profiling assay, the DASL assay (cDNA-mediated annealing, selection, extension, and ligation) based on the same universal Sentrix arrays used for SNP genotyping (Fan *et al.*, 2004). In this assay, RNA is first converted to cDNA by random priming; query oligonucleotides are designed for specific mRNA sequences in a way similar to the SNP assay design (Fig. 3, step 2) and bind to the corresponding cDNA and become extended and ligated enzymatically. The ligated products are then amplified and labeled fluorescently during PCR and are detected by hybridization to the universal array. The hybridization intensity is proportional to the original mRNA abundance in the sample and therefore represents the expression level of the targeted transcripts. The DASL assay protocol is identical to the GoldenGate assay described earlier (Fig. 3), except for the cDNA synthesis step. Specifically, a 20- μ l reverse transcription reaction containing a reaction mix (MMC; Illumina) and total RNA (up to 1 μ g) is incubated at room temperature for 10 min and then at 42° for 1 h.

The DASL assay multiplexes to over 1500 sequence targets, e.g., 500 genes at three probes per gene or 1500 genes at one probe per gene. It can measure gene expression level accurately across a dynamic range of 2.5–3 logs and resolve a less than twofold difference in expression between samples using 100 ng total RNA (Bibikova *et al.*, 2004a; Fan *et al.*, 2004). It fills the gap between two existing RNA profiling technologies: quantitative RT-PCR (high sample throughput/low gene content) and high-density oligonucleotide

or cDNA microarray technology (high gene content/low sample throughput), with medium to high sample/gene throughput and high sensitivity.

Using the DASL assay, highly reproducible tissue- and cancer-specific gene expression profiles have been obtained with partially degraded RNAs isolated from formalin-fixed, paraffin-embedded tissues that had been stored from 1 to over 10 years (Bibikova *et al.*, 2004a,b) (Fig. 5). The method can tolerate RNA degradation in these fixed tissues because it uses a random priming procedure for cDNA synthesis. In general, RNA is extracted from four to five 5- μ m tissue sections using the High Pure RNA Paraffin kit (Roche), yielding 0.3–15 μ g total RNA. For each FFPE sample, 200 ng of total RNA is converted to cDNA and processed using the DASL assay.

Due to its high sequence specificity, the DASL assay can be used for splice variant detection (Yeakley *et al.*, 2002), as well as allele-specific expression profiling (i.e., measuring the relative abundance of allelic

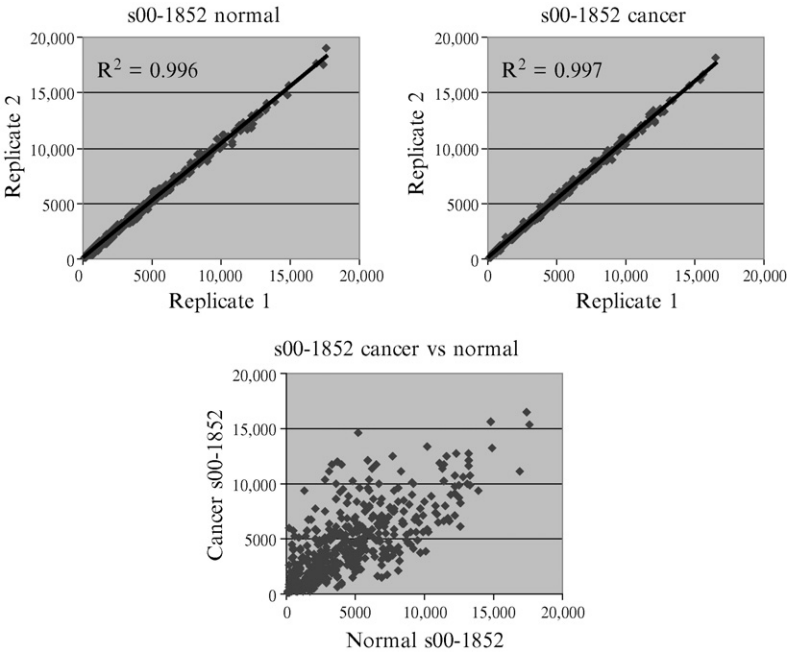


FIG. 5. Gene expression profiling in FFPE samples using the DASL assay. Highly reproducible gene expression profiles are obtained with RNAs isolated from normal and tumor FFPE tissue of the same patient (top panels). Differential expression (bottom panel) can be identified readily between normal and cancer samples.

transcripts) (Fan *et al.*, 2003). Differential allelic expression is relatively common, which occurs not only in imprinted genes, but also in non-imprinted autosomal genes (Knight, 2004; Pastinen and Hudson, 2004; Pastinen *et al.*, 2004; Yan *et al.*, 2002). We have developed genotyping assays for over 2000 SNPs derived from the coding region or 5'- and 3'-UTRs of more than 700 cancer-related genes (Fan *et al.*, unpublished data). Assay probes targeting specific SNPs are designed the same way as the standard GoldenGate SNP design, except that each assay target is chosen from the sense strand of mRNA and within one exon. Thus, they can be used to interrogate both genomic DNA (for genomic heterozygosity) and RNA (for relative abundance of allelic transcripts at the heterozygous loci). Imbalanced allelic expression is detected when the allele ratio in RNA (cDNA) differs from the corresponding ratio in genomic DNA (which is expected to be 1:1).

Conclusion

The GoldenGate genotyping platform combines a highly multiplexed genotyping assay with highly flexible Sentrix arrays to deliver high data quality and throughput. One of the most important features of the GoldenGate genotyping assay is that it genotypes directly on the genomic DNA and does not require prior PCR amplification of the genotyping target. It has been used in many genetic studies, including the International HapMap Project (The International HapMap Consortium, 2003), and has generated a large amount of genotyping data with an exceptional call accuracy and call rate. The scalability of the system allows the user to perform either small pilot studies or large-scale SNP genotyping association studies under the same system.

The DASL assay is particularly useful for profiling the expression of hundreds or thousands of genes in degraded RNAs, including RNAs from formalin-fixed, paraffin-embedded tissues. Formalin-fixed archival tissues represent an invaluable resource for genetic analysis and are the most widely available materials for which patient outcomes are already known. The ability to perform genetic analysis in these samples will enable both prospective and retrospective studies, facilitating research in correlating genetic profiles with clinical parameters.

In summary, BeadArray technology offers a complete solution to high-throughput genomic analysis, including SNP genotyping, LOH and genome amplification detection, DNA methylation profiling, and gene expression profiling, including allele-specific expression profiling and mRNA isoform profiling. It can be used to generate comprehensive genetic, epigenetic, and gene expression profiles for any biological system.

Acknowledgments

We thank John Stuelpnagel, Nicky Espinosa, and Melanie Smith for critical reading of this chapter and constructive suggestions. We are indebted to the many other scientists and engineers at Illumina whose dedicated work has created the platforms and assays reviewed here.

References

- Barker, D. L., Hansen, M. S., Faruqi, A. F., Giannola, D., Irsula, O. R., Lasken, R. S., Latterich, M., Makarov, V., Oliphant, A., Pinter, J. H., Shen, R., Sleptsova, I., Ziehler, W., and Lai, E. (2004). Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res.* **14**, 901–907.
- Barker, D. L., Theriault, D., Che, D., Dickinson, T., Shen, R., and Kain, R. (2003). Self-assembled random arrays: High-performance imaging and genomics applications on a high-density microarray platform. *Proc. SPIE* **4966**, 1–11.
- Bibikova, M., Talantov, D., Chudin, E., Yeakley, J. M., Chen, J., Doucet, D., Wickham, E., Atkins, D., Barker, D., Chee, M., Wang, Y., and Fan, J. B. (2004a). Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. *Am. J. Pathol.* **165**, 1799–1807.
- Bibikova, M., Yeakley, J. M., Chudin, E., Chen, J., Wickham, E., Wang-Rodriguez, J., and Fan, J. B. (2004b). Gene expression profiles in formalin-fixed, paraffin-embedded tissues obtained with a novel assay for microarray analysis. *Clin. Chem.* **50**, 2384–2386.
- Botstein, D., and Risch, N. (2003). Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33** (Suppl.), 228–237.
- Braeckmans, K., De Smedt, S. C., Leblans, M., Pauwels, R., and Demeester, J. (2002). Encoding microcarriers: Present and future technologies. *Nat. Rev. Drug Discov.* **1**, 447–456.
- Chan, W. C., Maxwell, D. J., Gao, X., Bailey, R. E., Han, M., and Nie, S. (2002). Luminescent quantum dots for multiplexed biological detection and imaging. *Curr. Opin. Biotechnol.* **13**, 40–46.
- Chen, J., Iannone, M. A., Li, M. S., Taylor, J. D., Rivers, P., Nelsen, A. J., Slentz-Kesler, K. A., Roses, A., and Weiner, M. P. (2000). A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res.* **10**, 549–557.
- Fan, J. B., Chen, X., Halushka, M. K., Berno, A., Huang, X., Ryder, T., Lipshutz, R. J., Lockhart, D. J., and Chakravarti, A. (2000). Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* **10**, 853–860.
- Fan, J. B., Oliphant, A., Shen, R., Kermani, B. G., Garcia, F., Gunderson, K. L., Hansen, M., Steemers, F., Butler, S. L., Deloukas, P., Galver, L., Hunt, S., McBride, C., Bibikova, M., Rubano, T., Chen, J., Wickham, E., Doucet, D., Chang, W., Campbell, D., Zhang, B., Kruglyak, S., Bentley, D., Haas, J., Rigault, P., Zhou, L., Stuelpnagel, J., and Chee, M. S. (2003). Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 69–78.
- Fan, J. B., Yeakley, J. M., Bibikova, M., Chudin, E., Wickham, E., Chen, J., Doucet, D., Rigault, P., Zhang, B., Shen, R., McBride, C., Li, H. R., Fu, X. D., Oliphant, A., Barker, D. L., and Chee, M. S. (2004). A versatile assay for high-throughput gene expression profiling on universal array matrices. *Genome Res.* **14**, 878–885.
- Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773.

- Fulton, R. J., McDade, R. L., Smith, P. L., Kienker, L. J., and Kettman, J. R., Jr. (1997). Advanced multiplexed analysis with the FlowMetrix system. *Clin. Chem.* **43**, 1749–1756.
- Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G., and Barany, F. (1999). Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* **292**, 251–262.
- Gunderson, K. L., Kruglyak, S., Graige, M. S., Garcia, F., Kermani, B. G., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J., Doucet, D., Milewski, M., Yang, R., Siegmund, C., Haas, J., Zhou, L., Oliphant, A., Fan, J. B., Barnard, S., and Chee, M. S. (2004). Decoding randomly ordered DNA arrays. *Genome Res.* **14**, 870–877.
- Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G., and Chee, M. S. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genet.* **37**, 549–554.
- Han, M., Gao, X., Su, J. Z., and Nie, S. (2001). Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nat. Biotechnol.* **19**, 631–635.
- Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A., and Cox, D. R. (2005). Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079.
- Holloway, A. J., van Laar, R. K., Tothill, R. W., and Bowtell, D. D. (2002). Options available—from start to finish—for obtaining data from DNA microarrays II. *Nat. Genet.* **32**(Suppl.), 481–489.
- Iannone, M. A., Taylor, J. D., Chen, J., Li, M. S., Rivers, P., SlentzKesler, K. A., and Weiner, M. P. (2000). Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. *Cytometry* **39**, 131–140.
- Knight, J. C. (2004). Allele-specific gene expression uncovered. *Trends Genet.* **20**, 113–116.
- Kuhn, K., Baker, S. C., Chudin, E., Lieu, M. H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T. K., and Chee, M. S. (2004). A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res.* **14**, 2347–2356.
- Kwok, P. Y., and Chen, X. (2003). Detection of single nucleotide polymorphisms. *Curr. Issues Mol. Biol.* **5**, 43–60.
- Lockhart, D. J., and Trulson, M. O. (2001). Multiplex metallica. *Nat. Biotechnol.* **19**, 1122–1123.
- Michael, K. L., Taylor, L. C., Schultz, S. L., and Walt, D. R. (1998). Randomly ordered addressable high-density optical sensor arrays. *Anal. Chem.* **70**, 1242–1248.
- Nicewarner-Pena, S. R., Freeman, R. G., Reiss, B. D., He, L., Pena, D. J., Walton, I. D., Cromer, R., Keating, C. D., and Natan, M. J. (2001). Submicrometer metallic barcodes. *Science* **294**, 137–141.
- Oliphant, A., Barker, D. L., Stuelpnagel, J. R., and Chee, M. S. (2002). BeadArray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* (Suppl.), 56–58, 60–61.
- Pastinen, T., and Hudson, T. J. (2004). Cis-acting regulatory variation in the human genome. *Science* **306**, 647–650.
- Pastinen, T., Sladek, R., Gurd, S., Sammak, A., Ge, B., Lepage, P., Lavergne, K., Villeneuve, A., Gaudin, T., Brandstrom, H., Beck, A., Verner, A., Kingsley, J., Harmsen, E., Labuda, D., Morgan, K., Vohl, M. C., Naumova, A. K., Sinnott, D., and Hudson, T. J. (2004). A survey of genetic and epigenetic variation affecting human gene expression. *Physiol. Genom.* **16**, 184–193.
- Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.
- Shen, R., Fan, J. B., Campbell, D., Chang, W., Chen, J., Doucet, D., Yeakley, J., Bibikova, M., Wickham Garcia, E., McBride, C., Steemers, F., Garcia, F., Kermani, B. G., Gunderson, K., and Oliphant, A. (2005). High-throughput SNP genotyping on universal bead arrays. *Mutat. Res.* **573**, 70–82.

- Syvanen, A. C. (2005). Toward genome-wide SNP genotyping. *Nat. Genet.* **37**(Suppl.), 5–10.
- The International HapMap Consortium (2003). The International HapMap Project. *Nature* **426**, 789–796.
- Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., and Kinzler, K. W. (2002). Allelic variation in human gene expression. *Science* **297**, 1143.
- Yeakley, J. M., Fan, J. B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M. S., and Fu, X. D. (2002). Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.* **20**, 353–358.

[4] Microarray Oligonucleotide Probes

By DAVID P. KREIL, ROSLIN R. RUSSELL, and STEVEN RUSSELL

Abstract

Oligonucleotide probes are increasingly the method of choice for many modern DNA microarray applications. They provide higher target specificity, probe selection gives improved experimental control of hybridization properties, and targeting of specific gene subsequences allows better discrimination of highly similar targets such as splice variants or gene families. Only recently has there been substantial progress in dealing with the complexities of probe set design and probe-specific signal interpretation. After a discussion of advantages and disadvantages of oligonucleotide probes in comparison to amplicons, this chapter focuses on recent advances and remaining key challenges in probe design and computational data analysis for spotted and *in situ*-synthesized oligonucleotide microarray technologies. Both experimental questions and computational aspects are addressed. Experimental issues discussed include the choice of an optimal number of probes per target and probe lengths and their influence on bias and random measurement noise, effects of different probe or substrate modifications, and laboratory protocols on signal specificity and sensitivity. Computational topics include practical considerations and a case study in probe sequence design, the exploitation of probing multiple target regions, and the modeling of probe sequence-specific signals. The current state of the art of the field is examined, and principled thermodynamic probe design criteria are proposed that are based on the free energy of the probe–target complex at the hybridization temperature rather than its melting temperature. Finally, this chapter notes and discusses an emerging trend in recent computational work toward a focus on signal interpretation rather than probe sequence design.