

# Chapter 6

## Genotyping Arrays

Michael J. Lodes, Dominic Suci, David Danley, and Andrew McShea

**Abstract** Although the most common use of DNA microarrays is gene expression profiling, microarrays are also used for many other applications, including genotyping, resequencing, SNP analysis, and DNA methylation assays. Here we describe genotyping arrays for Influenza A subtype identification and for upper respiratory pathogen diagnostics using standard hybridization techniques and we also describe resequencing, SNP, and methylation assays using an enzyme-based strategy [25, 26].

### 6.1 Pathogen Identification

The need for laboratory assays that rapidly identify infectious diseases is substantiated by a number of government initiatives for their development. The Epidemic Outbreak Surveillance (EOS) program was initiated by the U.S. Air Force as an Advanced Concept Technology Demonstration (ACTD) to create a diagnostic assay that would identify 10 to 20 viruses and bacteria that are associated with upper respiratory infections (URI). The Centers for Disease Control has also established its Laboratory Influenza Test program to develop an improved diagnostic assay(s) that detects seasonal flu and novel influenza A viruses.

As with the EOS program, the CDC recognizes the importance of being able to distinguish between flu and other URI infections. A device capable of identifying this combination of bacteria and viruses must be able to overcome complexities associated with potentially highly variable genomes, which is especially relevant

---

M.J. Lodes, D. Suci, D. Danley, and A. McShea  
CombiMatrix Corporation, 6500 Harbour Heights Parkway, Mukilteo, WA 98275  
e-mails: CombiMatrix Corp., mlodes@combimatrix.com  
CombiMatrix Corp., dsuci@combimatrix.com  
CombiMatrix Corp., ddanley@combimatrix.com  
CombiMatrix Corp., amcshea@combimatrix.com

to influenza A subtyping. To date, there are 16 identified hemagglutinin [1] and 9 neuraminidase subtypes for Influenza A. This RNA virus has a negative strand, segmented genome, and can infect a broad range of animals including humans. Identification of a virus subtype is typically by serological or molecular identification of the subtype of viral hemagglutinin (HA) and neuraminidase (NA) genes. Viruses with any combination of the hemagglutinin and neuraminidase subtypes can infect aquatic birds whereas fewer subtypes have been found to infect humans. However, interspecies transmission can occur after recombination or mixing of subtypes in birds or pigs [2, 16, 65]. In addition, new human strains of virus can arise by reassortment to accomplish antigenic shift when two or more subtypes infect the same host [3, 4].

Identification of influenza subtypes is routinely accomplished with viral detection (cell culture) and serological techniques such as complement fixation, hemagglutination, hemagglutination inhibition assays, and immunofluorescence methods [51-8]. Traditional methods are generally effective, but involve labor-intensive protocols and highly trained personnel. Because of their speed, specificity and sensitivity, genomic assays are ideal for complementing serological assays for identifying the genotype of an unknown specimen, especially in cases where antigenic tests are not specific enough to differentiate closely related groups [9-14].

Reverse transcription-polymerase chain reaction (RT-PCR) is widely used for virus identification [15-17]. However, a positive amplification can be verified only by subsequent assays to elaborate sequence information. By overcoming this limitation, microarrays and biosensors have become valuable tools for viral discovery, detection, genotyping, and sequencing [6, 9, 10, 13, 17-26].

Although traditional assays for pathogen detection and typing represent the gold standard, they alone cannot meet the future needs of rapid, sensitive, specific, and simple methods. For example, although immunological methods are excellent for determining influenza subtypes, they do not give detailed genetic information or information when antigenic shifts occur. RT-PCR techniques depend on specific primers, which may fail when corresponding viral sequences mutate.

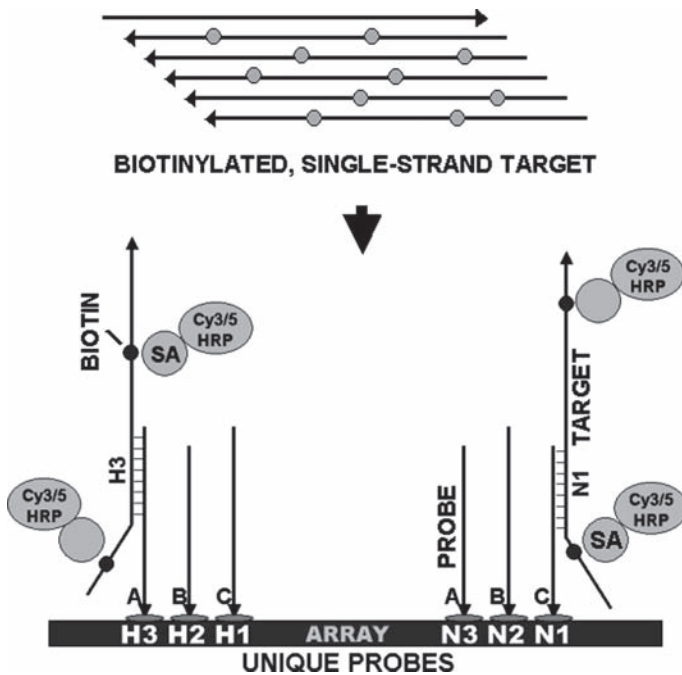
Microarrays offer an excellent solution as a downstream assay to PCR amplification. Semiconductor-based oligonucleotide array technology can be used with fluorescent labels and traditional optical scanning devices or used as a biodetector using electrochemical techniques for analysis [27]. This platform is extremely flexible, allowing array designs to be rapidly and easily modified and synthesized, and thus permitting oligonucleotides of interest to be tested empirically. In addition, the ability to use electrochemical detection with semiconductor microchips eliminates the need for expensive optical scanning equipment [27-29]. The use of an endpoint measurement on the PCR reaction, such as with a real-time PCR system, when preparing the target sample that will be hybridized to the array can be used as a decision point to "gate" the choice of samples to be run (i.e., assays are only run if the PCR reaction is positive).

Two genotyping arrays have been developed: an influenza A subtyping array that contains specific probes for each of 16 hemagglutinin (HA) subtypes and 9 neuraminidase (NA) subtypes and also an upper respiratory pathogen diagnostic array that can be used to detect both bacteria and viruses that produce upper res-

piratory clinical symptoms. These arrays can be hybridized with target generated from all 16 HA and 9 NA subtypes or from upper respiratory tract pathogens by a one-tube RT-PCR amplification of virus RNA or bacterial DNA.

The arrays were developed with nonoverlapping probes with similar annealing stabilities that were generated from sequence databases. For the influenza subtyping chip, subtype-specific probes were selected from a pool of over 23,000 HA and 15,000 NA sequences and then compared to the database to ensure that each probe was unique to the respective subtype and would hybridize to the maximum number of variant sequences. Subtype-specific probes were also subgrouped to give a finer detail for follow-up analysis, such as cluster analysis and sequence reconstruction. The probes for the upper respiratory arrays were developed in a similar manner and then cross-checked against databases to minimize the possibility of cross-hybridization of target and background from host genomic DNA.

The arrays were validated by hybridizing target (see Fig. 6.1) that was generated from RNA or DNA from all HA and NA subtypes on the influenza chip and all viruses and bacteria on the upper respiratory chip, which includes influenza A and B,

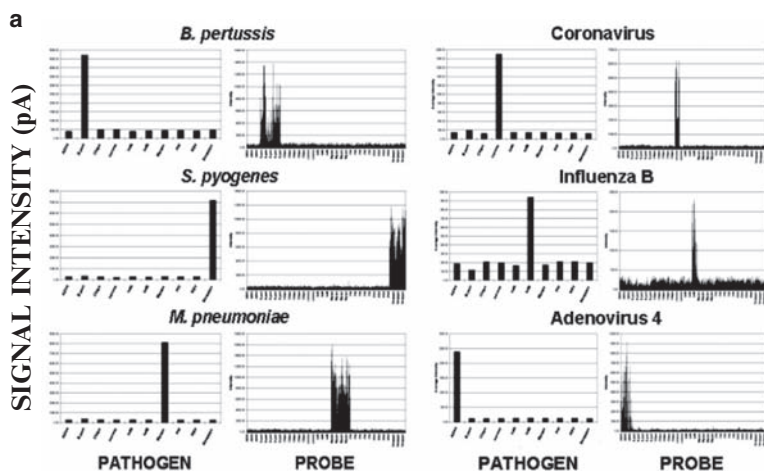


**Fig. 6.1** Diagram of target amplification and hybridization strategy for the identification of influenza A subtype H3N1. Briefly, a one-tube, three-stage RT-PCR reaction results in single-strand cDNA that is first amplified and then copied to produce biotin-labeled, single-strand target. Target is then hybridized to the array for 1 h and then labeled with either Streptavidin-Cy-5, fluorescent scanning, or with streptavidin-HRP, for electrochemical detection

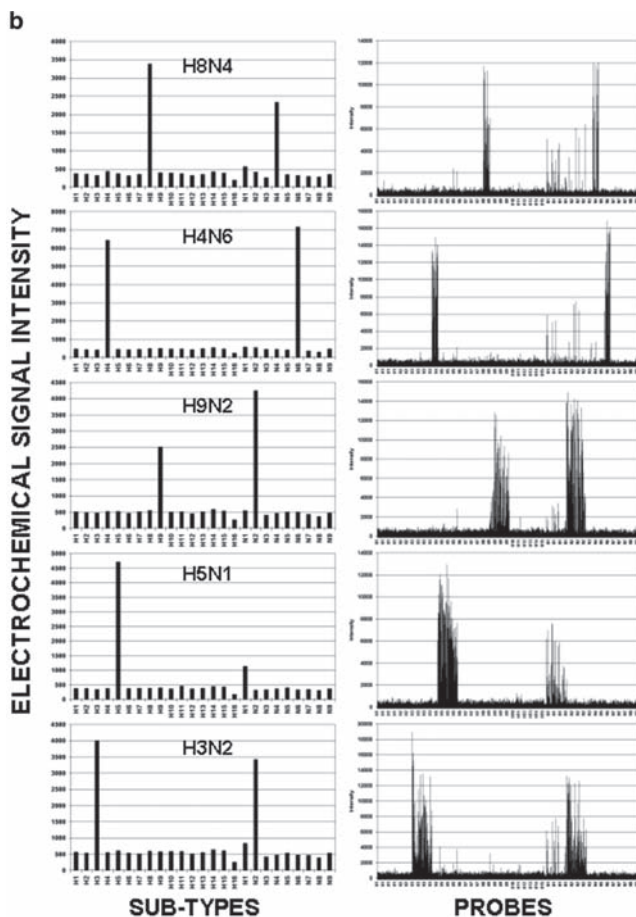
parainfluenza, adenovirus 4, respiratory syncytial virus, coronavirus, *Streptococcus pyogenes*, *Bordetella pertussis*, *Chlamydia pneumoniae*, and *Mycoplasma pneumonia* (see Fig. 6.2a,b). The target sample preparation system used with these arrays is similar to standard RT-PCR-based methods, except that it uses a very redundant/consensus priming system that maximizes the chance that novel strains of influenza will amplify, and thus, minimize false negative results.

Primers and probes for bacteria were developed from conserved genes and then compared to genomes of related organisms. The arrays contain multiple probes that correspond to key distinguishing elements of each organism or subtype. The combination of assay speed (approximately 1 hour hybridization), array sectoring which would allow multiple assays on one array, the potential to strip and reuse the chip up to five times (for conventional hybridizations), and the adaptability to inexpensive electrochemical scanning devices make these arrays a superb adjunct to real-time PCR by supporting multiplex assays and analyses in a single PCR tube.

Rapid identification of upper respiratory pathogens followed by HA and NA subtyping of influenza A viruses, will significantly decrease the time and cost for the identification of potential lethal virus and bacterial strains and lead to better treatment and management of infections. Microarray and biosensor technologies show great promise for virus detection and genotyping and are needed for rapid vaccine development, environmental screening, and the detection of bioterrorism agents [14, 15, 20, 30, 31].



**Fig. 6.2a** Results of upper respiratory pathogen identification using electrochemical detection. High signal intensities indicate the bacterial pathogen (left panels) and viral pathogen (right panels) identity. Probe signal intensities are shown to the right of the typing panels. Identified bacteria include: *Bordetella pertussis*, *Streptococcus pyogenes*, and *Mycoplasma pneumoniae*; and identified viruses include: Coronavirus, Influenza B, and Adenovirus 4 (see [26], PLoS ONE for details)



**Fig. 6.2b** Results of influenza A subtyping using electrochemical detection. High signal intensities indicate the hemagglutinin (H) and neuraminidase (N) subtypes. Subtype identities are shown in the left panels and probe signal intensities are shown in right panels

## 6.2 Genotyping Assay

Targets were labeled by biotin incorporation during RT-PCR. Briefly, 25  $\mu$ l reaction mixes included 12.5  $\mu$ l of reaction buffer (Invitrogen; SuperScript III One-Step RT-PCR kit with Platinum Taq), 2  $\mu$ l of 5 mM  $MgSO_4$ , 0.7  $\mu$ l of 0.4 mM biotin-14-dCTP (Invitrogen), 2  $\mu$ l primer pool (IDT, Coralville, IA), 0.5  $\mu$ l enzyme mix, 2  $\mu$ l RNA or DNA sample (diluted in a solution containing 40.0  $\mu$ g yeast tRNA (Invitrogen), and 9.6  $\mu$ g BSA (NEB) per 1.0 ml of  $dH_2O$ ), and 4.3  $\mu$ l  $dH_2O$ . Thermal cycling parameters were: (50°C – 30 min)  $\times$  1 cycle; (94°C – 4 min)  $\times$  1 cycle;

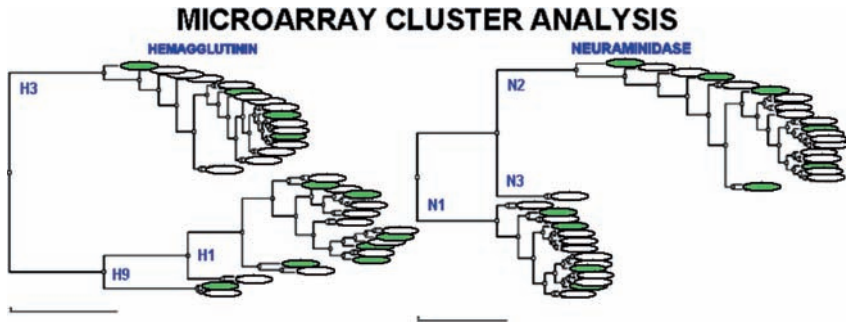
(94°C – 30 s, 56°C – 45 s, 72°C – 45 s) × 40 cycles; (94°C – 30 s, 68°C – 60 s) × 30 cycles and (72°C – 5 min) × 1 cycle. Primers were designed so that the forward primer  $T_m$  was approximately 50°C and the reverse primer  $T_m$  was approximately 65°C. Influenza subtyping was accomplished with a universal forward primer with a 5' tag and thus the RT stage of amplification was set to 42°C instead of 50°C (5'CTATAGGAGCAAAGCAGG). Amplification of target was confirmed by visualizing 3.0 µl of each reaction product on 6% polyacrylamide gels (Invitrogen) and staining with SYBR Green I (Molecular Probes, Invitrogen).

Initially, microarrays were prehybridized for 30 min at 45°C in 50 µl of a solution consisting of 5 ml of 2 × hybridization solution (see below), 1 ml of 50 × Denhardt's solution (Sigma), and 0.5 ml of 1% SDS (Sigma). For hybridization (see Fig. 6.1), PCR reactions from primer pools were combined and mixed 1:1 with 2 × hybridization buffer, which consisted of 6 ml of 20 × SSPE (Ambion, Austin, TX), 0.1 ml of 10% Tween 20 (Sigma), 0.56 ml of 0.5 M EDTA (Ambion), 0.5 ml of 1% SDS (Sigma), and 3.84 ml of dH<sub>2</sub>O (Ambion). Microarray hybridization chambers were filled (50 µl volume) and sealed with tape. The arrays were incubated for 1 h at 45°C with rotation in a hybridization oven (Fisher Scientific, Pittsburgh, PA) and washed for 5 min at 45°C with 3X SSPE with 0.05% Tween 20; twice with 2 × PBS with 0.1% Tween 20 (PBST); and then blocked for 5 min with 5 × PBS/Casein (BioFX Laboratories, Owing Mills, MD).

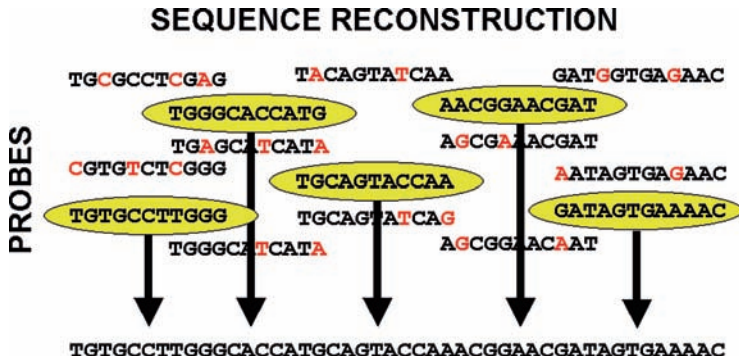
For labeling, microarrays were incubated for 30 min with ExtrAvidin Peroxidase (Sigma) diluted 1:1000 in BSA Peroxidase Stabilizer (BioFX). Arrays were washed twice with 2 × PBST, and twice with pH4 Conductivity Buffer Substrate (BioFX). TMB Conductivity 1 Component HRP Microwell Substrate (BioFX) was added to the array, and it was scanned immediately with an ElectraSense® microarray reader (CombiMatrix Corp.). This instrument measures µA at each of 12,544 electrodes on the array in 25 s and outputs data in picoamps to a simple text file that can be used to create a pseudoimage or can be transferred to and graphed with an Excel macro.

After hybridization to microarrays and data extraction by ECD, graphs of mean subtype intensity values can be used to predict the correct pathogen genotype or subtype. This information can next be broken down into subtype-subgroups. This grouped data can then be used to cluster samples into like groups with alignment software. This software treats each hybridization as an ordered list of intensities, where the value at each position corresponds to the intensity of a unique probe. A distance metric, such as correlation, allows for the determination of the difference between any two hybridization vectors. Next a similarity matrix is created containing the distances between all the hybridizations that are being compared. This similarity matrix can then be used with several clustering programs, such as the BioEdit Sequence Alignment Editor, as diagrammed in Fig. 6.3.

Sequence can also be reconstructed with array probe signal intensity data and then aligned with the most similar sequence in GenBank (Fig. 6.4). Briefly, sequence reconstruction software can take all the existing sequences in the database, align them against the chip probe sequences, and then create a probe profile for each sequence. For each hybridization that is analyzed, data are passed through these probe profiles, and the probe profiles with the highest scores are chosen. Next,



**Fig. 6.3** Cluster analysis: signal intensity data for each sample are compared and probe-to-probe correlations are determined. This data similarity matrix is then used to develop an output that shows the relationships of all samples to one another. Clusters have been grouped into the hemagglutinin subtypes H1, H3, and H9 and the neuraminidase subtypes N1, N2, and N3. Green ovals represent known reference samples to which unknown samples are compared



**Fig. 6.4** Sequence reconstruction from microarray data: multiple probes per pathogen, that represent the genetic diversity in genomic databases, are synthesized on an array and probed with an unknown target. Our proprietary software extracts the relevant probe signal and translates the associated probe sequence into a reconstructed sequence

sequence reconstruction proceeds within the small database containing the winning sequences. By limiting the dataset, resulting data have a lower background.

### 6.3 Single Nucleotide Polymorphisms/Resequencing

Oligonucleotide microarrays have also been developed for single nucleotide polymorphism (SNP) analysis and for resequencing of target DNAs. Over 10 million single nucleotide polymorphisms (SNP) have been estimated to occur in the human genome (International HapMap Project: [www.hapmap.org/](http://www.hapmap.org/)). Many of these

polymorphisms have been associated directly or indirectly with genetic diseases including Crohn's disease, ataxia telangiectasia, and Alzheimer's disease. Certain Crohn's disease patients, for example, have been shown to have mutations in one or more of several genes that are associated with increased susceptibility and disease behavior. These genes include the CARD15/NOD2 gene [32, 33], and the MDR1 gene [34]. Also, mutations in the ataxia telangiectasia mutated (ATM) gene have been associated with lymphoma [35] and ataxia telangiectasia, which is characterized by cerebellar and neuromotor degeneration and immune deficiency [36].

The ability to detect mutations in patient genomes allows for a specific diagnosis and therapy and also allows for the prediction of potential disease in patients with a family history of genetic disease. Several technologies have been utilized to detect known SNPs including hybridization on microarrays or in real-time PCR; enzymatic nucleotide extension, cleavage, or ligation; and mass spectroscopy (Reviewed in: [37–40]).

Oligonucleotide microarray-based hybridization analyses of SNPs have been used to screen for both previously characterized sequence variants and for the discovery of new sequence variants (reviewed in [41]). This approach uses either a gain of signal approach, where probes are complementary to sequence changes of interest, and measures gain of hybridization signal to these probes relative to reference samples; or loss of signal, which analyzes loss of hybridization signal to perfect match probes that are complementary to wild-type sequence [41].

Potential difficulties with this approach include the ability to detect heterozygous base changes versus homozygous mutations; intra- and intermolecular structures, such as hairpin and G-rich sequence; and G/C-rich versus A/T-rich sequence. In addition, either suboptimal hybridization conditions for G/C-rich sequence must be used to detect an A/T-rich sequence, or A/T-rich probes must be increased in length to equalize hybridization conditions, thus reducing one's ability to detect changes in the A/T-rich sequence [42]. Hybridization analyses suffer limitations in detecting known mutations and have demonstrated a high accuracy on only 65% of the DNA surveyed due to a high hybridization stringency [43].

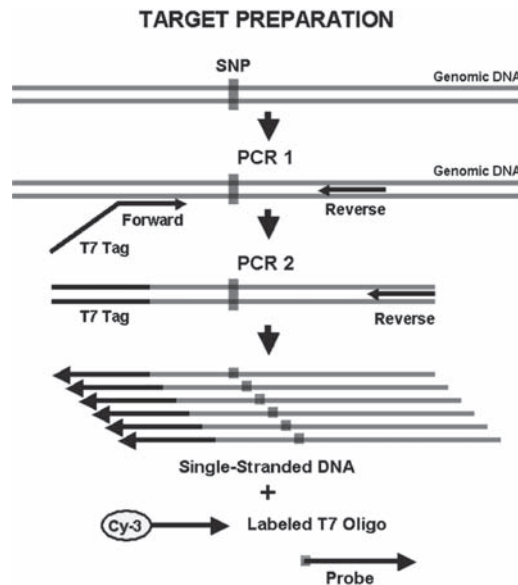
Enzymatic procedures have also been used for mismatch discrimination. These approaches include primer extension or minisequencing, and ligation of probes to sequence specific primers, using the genomic sequence as a hybridization template [44–47]. Primer extension, or minisequencing, involves the extension, on single-stranded, amplified genomic DNA, of a specific primer in the presence of polymerase and either fluorescent ddNTPs or 1 ddNTP and 3 dNTPs, and detection with gel or capillary sequencing or MALDI/TOF, respectively [37]. Ligation reactions usually require two adjacent primers to anneal to a genome-derived target. The upstream primer usually contains a label on the 5' end and the 3' nucleotide is designed to be opposite the SNP of interest. When the 3' nucleotide forms a perfect match with the target, the primer, with label, is covalently attached by ligase to the downstream primer.

Detection is by fluorescent display on a microarray or by MALDI/TOF [37, 48–51]. One disadvantage in this procedure is the expense of using labeled specific primers for SNPs being screened.

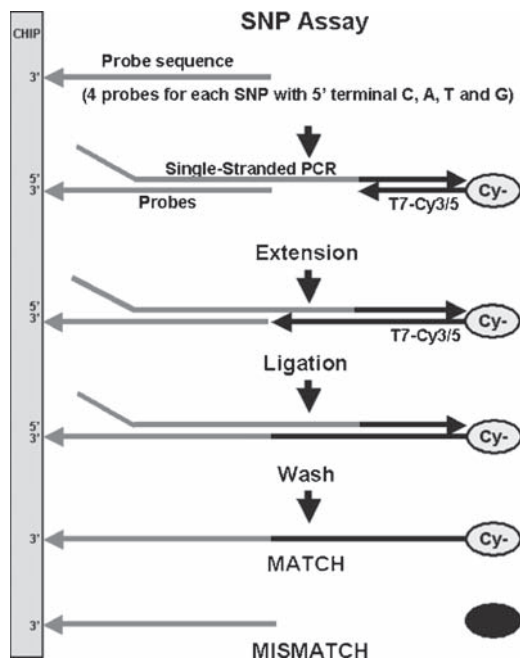


A microchip-based multiplex SNP assay that combines the sensitivity and specificity of ligation with the cost-effective strategy of using a labeled common oligonucleotide that is extended to the site of the match/mismatch is described next. This system can also be used for resequencing DNA by interrogating each nucleotide of interest. Briefly, a target sequence is amplified from the patient genomes of interest, with two specific primers, one containing a common tag for reamplification and detection (See Fig. 6.5). The common tag, which is added during amplification, provides the template for primer extension and an antisense sequence for labeled primer hybridization. The procedure consists of an on-chip hybridization of the combined single-stranded PCR product and labeled common primer, followed by a single-step extension/ligation reaction that can be accomplished with a DNA ligase such as *E. coli* DNA ligase and a polymerase such as Taq Stoffel fragment or reverse transcriptase (Fig. 6.6). The selection of appropriate enzymes for this combined reaction is critical. The polymerase must not have strand displacement or exonuclease activity and the ligase must be able to discriminate a mismatch. A final stringent wash step with NaOH removes unligated label and allows viewing of the SNP (Fig. 6.7) or sequence of interest (Fig. 6.8).

In many situations, identification of a pathogen is not sufficient and a specific gene sequence is required. For example, genotype Z, the dominant avian H5N1



**Fig. 6.5** Diagram of target amplification for SNP analysis and resequencing chips. Briefly, the genomic DNA of interest is amplified with a pair of specific primers. The forward primer contains a tag sequence (T7) for labeling and primer extension with a Cy-5 or biotin labeled oligonucleotide. A second amplification with only the reverse primer results in single-stranded target for hybridization to the array probes

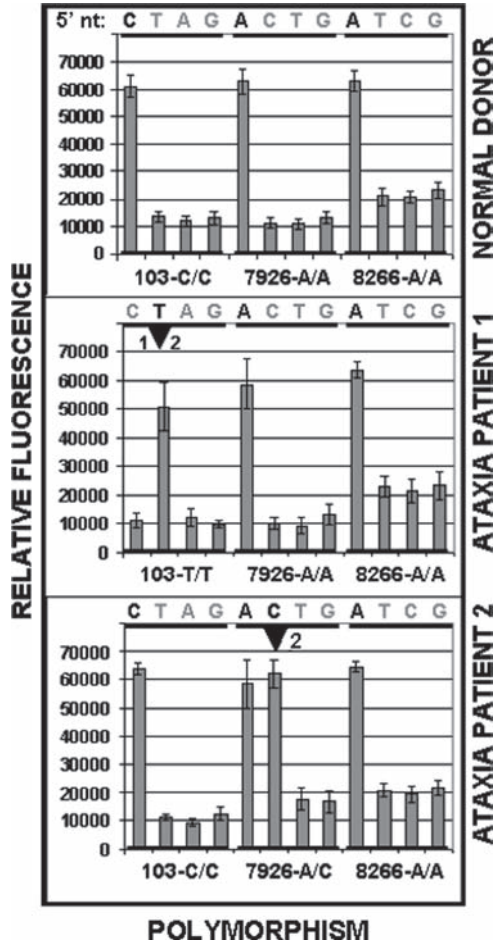


**Fig. 6.6** Diagram of a SNP assay (or resequencing assay). Single-stranded target (see Fig. 6.5) and labeled primer (T7-Cy3/5) are coincubated on the array. After annealing of target and labeled primer, a mixture of polymerase and ligase are added to extend the labeled primers and ligate to probes. The array is then washed with NaOH to remove all unligated label and then scanned for fluorescence (or current for electrochemical detection). Identification of SNPs or generation of sequence is accomplished by detection of signal associated with the correctly ligated primer as only one of four (or two of four in the case of a heterozygous SNP) probes will be ligated to the labeled extended primer

virus genotype currently circulating in Vietnam and Thailand, contains a mutation that is associated with resistance to amantadine and rimantadine [31, 52].

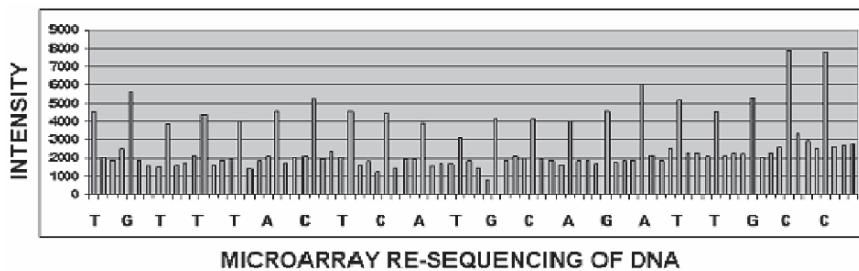
Antiviral therapies generally should be given within 48 hours of onset of illness to be effective against human influenza [31]. Thus rapid and specific identification of this subtype and accurate sequence information is crucial for proper treatment. Also, highly pathogenic strains of H5 and H7 influenza viruses can, in some cases, be distinguished from low pathogenic strains by sequencing the hemagglutinin gene, especially the area encoding the cleavage site [53].

With the enzyme-based SNP/resequencing assay described here, one is able to sequence approximately 500 or more nucleotides of genes of interest. This resequencing array and assay resulted in approximately 95% or more accurately called bases (998 of 1043 bases; [25]). Miscalls were predominately due to strong secondary structures, which can be predicted and avoided before the assay is carried out, and the ligated mismatches A/G, A/A, T/G, G/G, and T/T. These mismatches are generally more difficult to detect because of their low  $\Delta G$  values. For



**Fig. 6.7** Results of a SNP assay on Ataxia patient genomic DNA. Three potential SNPs in the human Ataxia-telangiectasia-mutated (ATM) gene were interrogated from two patient genomic DNA samples and one control normal donor sample. Results show the wild-type sequence (first bars) in the normal donor DNA and either homozygous (patient 1, arrowhead) or heterozygous (patient 2, arrowhead) mutations in the ATM patient DNA. Signal intensity, indicated by relative fluorescence, is shown at the left and the locations of polymorphisms within the ATM gene, are shown at the bottom

example, sequence errors resulting from A/G mismatches represented 1.4% of the total errors or 6% of the potential A/G mismatches (15 of 251). Strong secondary structures (hairpins and palindromes) interfere with probe-target hybridization and result in a reduced signal. These sequencing arrays should contain either a consensus subtype sequence or a known subtype sequence that lacks a high degree of secondary structure. Replicate probes for each base of sequence should reduce artifacts due to difficult mismatches by averaging out mismatch signal.



**Fig. 6.8** Example of a segment of DNA sequence from a resequencing array: this assay is essentially a SNP assay where each nucleotide in the sequence of interest is interrogated. Signal intensity is shown at the left and resulting sequence is shown below

Because microarray-based sequencing is based on probe-target hybridization, the target sequence cannot diverge significantly from the arrayed sequence. However, under nonstringent hybridization conditions, internal mismatches between probe and target sequence do not have as great an impact on hybridization and sequencing. This technique is best suited for resequencing, sequencing similar viruses such as seasonal quasispecies complexes, or for surveying mutations in an isolate over time.

### 6.3.1 Resequencing and SNP Assay

The oligonucleotide probes synthesized on the resequencing microarray chip were 5' phosphorylated with T4 polynucleotide kinase (PNK, New England Biolabs, Beverly, MA) for 30 min at 37°C. The array was then preblocked for 15 min at 45°C with 6 × SSPE containing 0.05% Tween-20 (SSPET), 2.0 mM EDTA, 5 × Denhardt's solution, and 0.05% SDS. Single-stranded target, with an antisense T7 tag added with the forward primer (5' TAATACGACTCACTATAGGAG CAAAAGCAGG) during PCR as shown in Fig. 6.5 (see Section 6.1 for details), was heated to 95°C for 10 min, placed on ice and T4 ligase buffer added to 1 × concentration. A 5' biotin-labeled T7 oligonucleotide (Integrated DNA Technologies, Inc., Coralville, IA) was added to a concentration of 1 μM to provide signal for detection and primer for extension. This solution was added to the chip array hybridization chamber and incubated at 45°C for 1 h.

After washing the array with 2 × PBS-0.5% Tween 20, 2 × PBS and 1 × *E. coli* ligase buffer, a mixture consisting of 1 × *E. coli* ligase buffer, 0.2 mM dNTP, 20 units each of AmpliTaq DNA polymerase, Stoffel fragment (Applied Biosystems, Foster City, CA), and *E. coli* ligase were added to the array and incubated at 37°C for 30 min. The array was washed twice for 2 min each with 0.1 N NaOH at room temperature, blocked and labeled as described in Section 6.1, and then scanned with ElectraSense (CombiMatrix Corp., Mukilteo, WA).

## 6.4 DNA Methylation

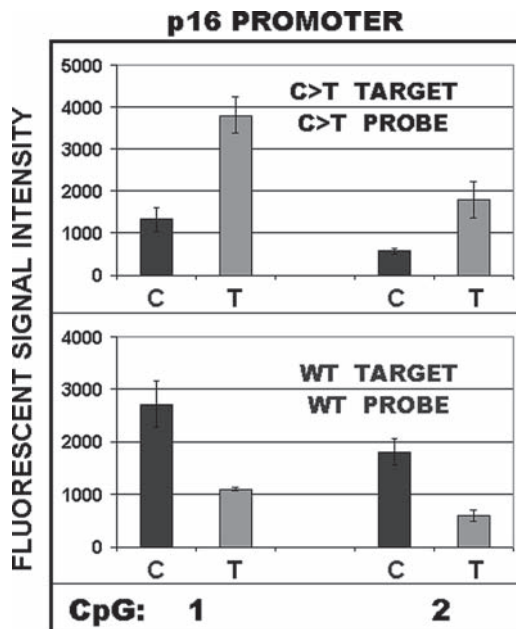
The methylation of DNA is a general mechanism for control of transcription in vertebrates. Cytosine residues in vertebrate DNA can be modified by the addition of methyl groups at the 5-carbon position. DNA is methylated specifically at the Cs that precede Gs in the DNA chain (CpG dinucleotides). This methylation is correlated with reduced transcriptional activity of genes that contain high frequencies of CpG dinucleotides in the vicinity of their promoters. Methylation inhibits transcription of these genes via the action of a protein, MeCP2, that specifically binds to methylated DNA and represses transcription [54]. The analysis of methylation patterns is also fundamental for the understanding of cell differentiation, X chromosome inactivation, regulation of developmental programming, aging processes, diseases, and cancer development [55–58].

Methylated DNA can be detected on oligonucleotide microarrays by using several techniques including methylation-specific restriction enzyme (MSRE) analysis [59], differential methylation hybridization (DMH) technique [60], and integrated analysis of methylation by isoschizomers [61]. One can also use standard hybridization to detect mismatches or the enzyme-based technique described in the SNP/resequencing section. For these methods, detection of methylated cytosines in genomic DNA involves PCR amplification of chemically modified DNA in which unmethylated cytosine residues have been converted to uracil by hydrolytic deamination, but methylated cytosine residues remain unconverted. Urea may improve efficiency of bisulphite-mediated sequencing of 5[prime]-methylcytosine in genomic DNA [62].

Briefly, genomic DNA is digested with endonucleases or sonicated to produce short fragments. Fragmented DNA is then denatured with sodium hydroxide and treated with sodium metabisulphite and hydroquinone and cycled at 55°C and 95°C in a thermocycler [63]. The bisulphite treated DNA is further processed by desalting, with a 3 M column, for example, and by elution in methanol, desiccation, and resuspension in water [64]. Finally the DNA is treated with sodium hydroxide to complete the conversion of unmethylated C residues to Ts. The conversion of unmethylated Cs and the nonconverted methylated Cs can be detected with SNP or resequencing arrays. Results of an enzyme-based assay used to detect methylated DNA are shown in [Fig. 6.9](#).

### 6.4.1 Methylation Assay

CpG methylation is accomplished by sonicating 2 ug of genomic DNA to product fragments of approximately 500 to 1000bp. Genomic DNA is diluted to 0.2ug/ul and sonicated at a power setting of 3 for 10s. Sonicated DNA is then purified with a Qiagen nucleotide removal column. The DNA is then denatured by the addition of 1/9 vol of freshly prepared 3M sodium hydroxide and incubation for 15 min



**Fig. 6.9** Illustration of a methylation assay to show the position of CpG dinucleotides. Random donor genomic DNA was sonicated and divided into two groups: wild-type (WT) and converted (C > T). For conversion, unmethylated cytosine residues are converted to uracil by hydrolytic deamination and methylated cytosine residues remain unconverted. The region of genomic DNA containing the CpG residues of interest was amplified from the wild-type and converted DNAs using the strategy illustrated in Fig. 6.5 and then hybridized to arrays designed to identify the wild-type Cytosine (C) residue or the converted Thymidine (T) residue. Fluorescence intensity is shown to the left and the CpG dinucleotides being assayed (1 or 2) are shown below. The top panel illustrates the results of hybridization of the converted target and the bottom panel illustrates results of hybridization of the wild-type target using the protocol described in Fig. 6.6

at 37°C. A 6.24M urea/2M sodium metabisulphite (4M bisulphite) solution is made by dissolving 7.5g of urea in 10ml of sterile distilled water, adding 7.6g of sodium metabisulphite (8.5g sodium bisulphate), adjusting the pH to 5 with 10M sodium hydroxide and adding sterile water to a final volume of 20ml. The urea/bisulphite solution and 10mM hydroquinone are then added to the denatured DNA to final concentrations of 5.36M, 3.44M, and 0.5mM, respectively.

The reaction is performed in a 0.5ml PCR tube overlaid with 100µl of mineral oil and cycled 20 times at 55°C for 15min followed by denaturation at 95°C for 30s [63]. The bisulphite-treated DNA is desalted with a 3M column: Add 9 parts NEN A (0.1M Tris 7.7, 1mM EDTA, and 10mM TEA), load onto a 3M EMPORE column and centrifuge into a 15ml tube and then wash by centrifugation with 10 parts NEN A. Repeat centrifugation until dry. Elute in 500ul NEN B

(50% MeOH/dH<sub>2</sub>O) and dry in a lyophilizer. Resuspend pellet in 90  $\mu$ l dH<sub>2</sub>O and add fresh 3 M NaOH to a concentration of 0.3 M (1:10) and incubate at 37°C for 15 min. Precipitate by adding NH<sub>4</sub>OAc, pH 7, to the 3 M column-purified material (3 M final) and precipitate with EtOH. Dry with a lyophilizer, resuspend in 100  $\mu$ l TE buffer, and store at -20°C until analyzed. PCR conditions to produce single-stranded target from methylated DNA and wild-type DNA are described in Section 6.3 as for resequencing arrays.

**Acknowledgments** We would like to thank Jodi Dalrymple and Marty Ross for their contributions to this study.

## References

1. Fouchier R.A., V. Munster, A. Wallensten, T.M. Bestebroer, S. Herfst, D. Smith, G.F. Rimmelzwaan, B. Olsen, and A.D. Osterhaus. (2005) Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J. Virol.* 79(5), 2814–2822.
2. Scholtissek, C., H. Burger, O. Kistner, and K.F. Shortridge. (1985) The nucleoprotein as a possible major factor in determining host specificity of influenza H3N2 viruses. *Virology.* 147(2), 287–294.
3. Mizuta, K., N. Katsushima, S. Ito, K. Sanjoh, T. Murata, C. Abiko, and S. Murayama. (2003) A rare appearance of influenza A(H1N2) as a reassortant in a community such as Yamagata where A(H1N1) and A(H3N2) co-circulate. *Microbiol. Immunol.* 47(5), 359–361.
4. Webby, R.J., and R.G. Webster. (2001) Emergence of influenza A viruses. *Phil. Trans. R. Soc. Lond. B356*, 1815–1826.
5. Allwinn, R., W. Preiser, H. Rabenau, S. Buxbaum, M. Sturmer, and H.W. Doerr. (2002) Laboratory diagnosis of influenza—Virology or serology? *Med. Microbiol. Immunol. (Berl)* 191(3–4), 157–160.
6. Amano, Y., and Q. Cheng. (2005) Detection of influenza virus: Traditional approaches and development of biosensors. *Anal. Bioanal. Chem.* 381(1), 156–184.
7. Palmer D.F., M.T. Coleman, W.R. Dowdle, and G.C. Schild. (1975) Advanced laboratory techniques for influenza diagnosis. Immunology Series No. 6. U.S. Department of Health, Education, and Welfare. p. 51–52.
8. Ueda, M., A. Maeda, N. Nakagawa, T. Kase, R. Kubota, H. Takakura, A. Ohshima, and Y. Okuno. (1998) Application of subtype-specific monoclonal antibodies for rapid detection and identification of influenza A and B viruses. *J. Clin. Microbiol.* 36(2), 340–344.
9. Ellis, J.S., and M.C. Zambon. (2002) Molecular diagnosis of influenza. *Rev. Med. Virol.* 12(6), 375–389.
10. Korimbocus J., N. Scaramozzino, B. Lacroix, J.M. Crance, D. Garin, and G. Vernet. (2005) DNA probe array for the simultaneous identification of herpes viruses, enteroviruses, and flaviviruses. *J. Clin. Microbiol.* 43(8), 3779–3787.
11. Schweiger, B., I. Zadow, and R. Heckler. (2002) Antigenic drift and variability of influenza viruses. *Med. Microbiol. Immunol. (Berl)*. 191(3–4), 133–138.
12. Taubenberger, J.K., and S.P. Layne. (2001) Diagnosis of influenza virus: Coming to grips with the molecular era. *Mol. Diagn.* 6(4), 291–305.
13. Wang, D., L. Coscoy, M. Zylberberg, P.C. Avila, H.A. Boushey, D. Ganem, and J.L. DeRisi. (2002) Microarray-based detection and genotyping of viral pathogens. *PNAS.* 99(4), 15687–15692.
14. Zou, S. (1997) A practical approach to genetic screening for influenza virus variants. *J. Clin. Microbiol.* 35(10), 2623–2627.

15. Adeyefa, C.A., K. Quayle, and J.W. McCauley. (1994) A rapid method for the analysis of influenza virus genes: Application to the reassortment of equine influenza virus genes. *Virus Res.* 32(3), 391–399.
16. Hoffmann, E., J. Stech, Y. Guan, R.G. Webster, and D.R. Perez. (2001) Universal primer set for the full-length amplification of all influenza A viruses. *Arch. Virol.* 146, 1–15.
17. Bodrossy, L., and A. Sessitsch. (2004) Oligonucleotide microarrays in microbial diagnosis. *Curr. Opin. Microbiol.* 7(3), 245–254.
17. Templeton, K.E., S.A. Scheltinga, M.F.C. Beersma, A.C.M. Kroes, and E.C.J. Claas. (2004) Rapid and sensitive method using multiplex real-time PCR for diagnosis of infections by influenza A and influenza B viruses, respiratory syncytial virus, and parainfluenza viruses 1, 2, 3, and 4. *J. Clin. Microbiol.* 42(4), 1564–1569.
18. Baxi, M.K., S. Baxi, A. Clavijo, K.M. Burton, and D. Deregt. (2006) Microarray-based detection and typing of foot-and-mouth disease virus. *Vet. J.* 172(3):473–481.
19. Ivshina, A.V., G.M. Vodeiko, V.A. Kuznetsov, D. Volokhov, R. Taffs, V.I. Chizhikov, R.A. Levandowski, and K.M. Chumakov. (2004) Mapping of genomic segments of influenza B virus strains by an oligonucleotide microarray method. *J. Clin. Microbiol.* 42(12), 5793–5801.
20. Kessler N., O. Ferraris, K. Palmer, W. Marsh, and A. Steel. (2004) Use of the DNA flow-thru chip, a three-dimensional biochip, for typing and subtyping of influenza viruses. *J. Clin. Microbiol.* 42(5), 2173–2185.
21. Li, J., S. Chen, and D.H. Evans. (2001) Typing and subtyping influenza virus using DNA microarrays and multiplex reverse transcriptase PCR. *J. Clin. Microbiol.* 39(2), 696–704.
22. Sengupta, S., K. Onodera, A. Lai, and U. Melcher. (2003) Molecular detection and identification of influenza viruses by oligonucleotide microarray hybridization. *J. Clin. Microbiol.* 41(10), 4542–4550.
23. Wang, D., A. Urisman, Y.-T. Liu, M. Springer, T.G. Ksiazek, D.D. Erdman, E.R. Mardis, M. Hickenbotham, V. Magrini, J. Eldred, J.P. Latreille, R.K. Wilson, D. Ganem, and J.L. DeRisi. (2003) Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol.* 1(2), 257–260.
24. Roth, S.B., J. Jalava, O. Ruuskanen, A. Ruohola, and S. Nikkari. (2004) Use of an oligonucleotide array for laboratory diagnosis of bacteria responsible for acute upper respiratory infections. *J. Clin. Microbiol.* 42 (9), 4268–4274.
25. Lodes, M.J., D. Suciú, M. Elliott, A.G. Stover, M. Ross, M. Caraballo, K. Dix, J. Crye, R.J. Webby, W.J. Lyon, D.L. Danley, and A. McShea. (2006) Use of semiconductor-based oligonucleotide microarrays for influenza A virus subtype identification and sequencing. *J. Clin. Microbiol.* 44(4), 1209–1218.
26. Lodes, M.J., D. Suciú, J.L. Wilmoth, M. Ross, K. Dix, K. Bernards, A.G. Stöver, M. Quintana, N. Iihoshi, W.J. Lyon, D.L. Danley, and A. McShea. (2007) Identification of upper respiratory tract pathogens using electrochemical detection on an oligonucleotide microarray. *PLoS ONE.* 2(9), e924.
27. Dill, K., D.D. Montgomery, A.L. Ghindilis, K.R. Schwarzkopf, S.R. Ragsdase, and A.V. Oleinikov. (2004) Immunoassays based on electrochemical detection using microelectrode arrays. *Biosens. Bioelectron.* 20, 736–742.
28. Ghindilis, A.L., Smith, M.W., Schwarzkopf, K.R., Roth, K., Peyvan, K., Munro, S., Lodes, M.J., Stover, A., Bernards, K., Dill, K., and McShea, A. (2007) CombiMatrix oligonucleotide arrays: Genotyping and gene expression assays employing electrochemical detection. *Biosens. Bioelectron.* 22(9–10):1853–1860.
29. Liu, R.H., M.J. Lodes, T. Nguyen, T. Siuda, M. Slota, H.S. Fuji, and A. McShea. (2006) Validation of a fully integrated microfluidic array device for influenza A subtype identification and sequencing. *Anal. Chem.* 2006 Jun 15; 78(12), 4184–4193.
30. Macken, C., H. Lu, J. Goodman, and L. Boykin. (2001) The value of a database in surveillance and vaccine selection. in *Options for the Control of Influenza IV.* A.D.M.E. Osterhaus, N. Cox, and A.W. Hampson (Eds.). Amsterdam: Elsevier Science, 2001, 103–106.



31. Yuen, K.Y. and S.S.Y. Wong. (2005) Human infection by avian influenza A H5N1. *Hong Kong Med. J.* 11(3), 189–199.
32. Helio, T., L. Halme, M. Lappalainen, H. Fodstad, P. Paavola-Sakki, U. Turunen, M. Farkkila, T. Krusius, and K. Kontula. (2003) CARD15/NOD2 gene variants are associated with familiarly occurring and complicated forms of Crohn's disease. *Gut.* 52(4), 558–562.
33. Newman, B., M.S. Silverberg, X. Gu, Q. Zhang, A. Lazaro, A.H. Steinhart, G.R. Greenberg, A.M. Griffiths, R.S. McLeod, Z. Cohen, M. Fernandez-Vina, C.I. Amos, and K. Siminovitch. (2004) CARD15 and HLA DRB1 alleles influence susceptibility and disease localization in Crohn's disease. *Am. J. Gastroenterol.* 99(2), 306–315.
34. Brant, S.R., C.I. Panhuysen, D. Nicolae, D.M. Reddy, D.K. Bonen, R. Karaliukas, L. Zhang, E. Swanson, L.W. Datta, T. Moran, G. Ravenhill, R.H. Duerr, J.P. Achkar, A.S. Karban, and J.H. Cho. (2003) MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. *Am. J. Hum. Genet.* 73(6), 1282–1292.
35. Fang, N.Y., T.C. Greiner, D.D. Weisenburger, W.C. Chan, J.M. Vose, L.M. Smith, J.O. Armitage, R.A. Mayer, B.L. Pike, F.S. Collins, and J.G. Hacia. (2003) Oligonucleotide microarrays demonstrate the highest frequency of ATM mutations in the mantle cell subtype of lymphoma. *PNAS.* 100(9), 5372–5377.
36. Hacia, J.G., B. Sun, N. Hunt, K. Edgemon, D. Mosbrook, C. Robbins, S.P.A. Fodor, D.A. Tagle, and F.S. Collins (1998) Strategies for mutational analysis of the large multiexon ATM gene using high-density oligonucleotide arrays. *Genome Res.* 8, 1245–1258.
37. Kirk, B.W., M. Feinsod, R. Favis, R.M. Kliman, and F. Barany, (2002) Single nucleotide polymorphism seeking long-term association with complex diseases. *Nucleic Acids Res.* 30(15), 3295–3311.
38. Kwok, P.-Y. (2001) Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet.* 2, 235–258.
39. Jenkins, S., and N. Gibson. (2002) High-throughput SNP genotyping. *Comp. Funct. Genom.* 3, 57–66.
40. Shi, M.M. (2001) Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin. Chem.* 47(2), 164–172.
41. Hacia, J.G. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics.* Supplement 21, 42–47.
42. Urakawa, H., S. El Fantroussi, H. Smidt, J.C. Smoot, E.H. Tribou, J.J. Kelly, P.A. Noble, and D.A. Stahl. (2003) Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays. *Appl. Environ. Microbiol.* 69(5), 2848–2856.
43. Patil, N., A.J. Berno, D.A. Hinds, W.A. Barrett, J.M. Doshi, C.R. Hacker, C.R. Kautzer, D.H. Lee, C. Marjoribanks, D.P. McDonough, B.T. Nguyen, M.C. Norris, J.B. Sheehan, N. Shen, D. Stern, R.P. Stokowski, D.J. Thomas, M.O. Trulson, K.R. Vyas, K.A. Frazer, S.P. Fodor, and D.R. Cox. (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science.* 294, 1719–1723.
44. Broude, N.E., T. Sano, C.L. Smith, and C.R. Cantor. (1994) Enhanced DNA sequencing by hybridization. *PNAS.* 91, 3072–3076.
45. Dubiley S, E. Kirillov, and A. Mirzabekov. (1999) Polymorphism analysis and gene detection by minisequencing on an array of gel-immobilized primers. *PNAS.* 27(18), e19.
46. O'Meara, D., A. Ahmadian, J. Odeberg, and J. Lundeberg. (2002) SNP typing by apyrase-mediated allele-specific primer extension on DNA microarrays. *NAR.* 30(15), e75.
47. Rickert, A.M., T.A. Borodina, E.J. Kuhn, H. Lehrach, and S. Sperling. (2004) Refinement of single-nucleotide polymorphism genotyping methods on human genomic DNA: Amplifluor allele-specific polymerase chain reaction versus ligation detection reaction-TaqMan. *Anal. Biochem.* 330(2), 288–297.
48. Zhong, X-b., R. Reynolds, J.R. Kidd, K.K. Kidd, R. Jenison, R.A. Marlar, and D.C. Ward. (2003) Single-nucleotide polymorphism genotyping on optical thin-film biosensor chips. *PNAS.* 100(20), 11559–11564.

49. Iannone, MA., J.D. Taylor, J. Chen, M.S. Li, P. Rivers, K.A. Slentz-Kesler, and M.P. Weiner. (2000) Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. *Cytometry*. 39(2), 131–140.
50. Chen, X., K.J. Livak, and P.-Y. Kwok. (1998) A homogeneous, ligase-mediated DNA diagnostic test. *Genome Res.* 8, 549–556.
51. Consolandi, C., A. Frosini, C. Pera, G.B. Ferrara, R. Bordoni, B., Castiglioni, E. Rizzi, A. Mezzelani, L.R. Bernardi, G. DeBellis, and C. Battaglia. 2004. Polymorphism analysis within the HLA-A locus by universal oligonucleotide array. *Hum. Mutat.* 24(5), 428–434.
52. Chen, H., G. Deng, Z. Li, G. Tian, Y. Li, P. Jiao, L. Zhang, Z. Liu, R.G. Webster, and K.Yu. (2004) The evolution of H5N1 influenza viruses in ducks in southern China. *PNAS.* 101(28), 10452–10457.
53. Lee, C-W., D.E. Swayne, J.A. Linares, D.A. Senne, and E.L. Suarez, (2005) H5N2 Avian influenza outbreak in Texas. *J. Virol.* 79(17), 11412–11421.
54. Cooper, G.M. (2000) *The Cell: A Molecular Approach*. Second edition, Boston University. Sinauer Associates, Sunderland, MA.
55. Schumacher A., P. Kapranov, Z. Kaminsky, J. Flanagan, A. Assadzadeh, P. Yau, C. Virtanen, N. Winegarten, J. Cheng, T. Gingeras, and A. Petronis. (2006) Microarray-based DNA methylation profiling: Technology and applications. *Nucleic Acids Res.* 34(2), 528–542.
56. Gebhard C., L. Schwarzfischer, T.H. Pham, E. Schilling, M. Klug, R. Andreesen, and M. Rehli. (2006) Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res.* 66(12), 6118–6128.
57. Lewin, J., A.O. Schmitt, P. Adorjan, T. Hildmann, and C. Piepenbrock. (2004) Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplicates. *Bioinformatics.* 20(17), 3005–3012.
58. Galm O., J.G. Herman, and S.B. Baylin. (2006) The fundamental role of epigenetics in hematopoietic malignancies. *Blood Rev.* 20(1), 1–13.
59. Adrien L.R., N.F. Schlecht, N. Kawachi, R.V. Smith, M. Brandwein-Gensler, A. Massimi, S. Chen, M.B. Prystowsky, G. Childs, and T.J. Belbin. (2006) Classification of DNA methylation patterns in tumor cell genomes using a CpG island microarray. *Cytogenet Genome Res.* 114(1), 16–23.
60. Yan P.S., C.M. Chen, H. Shi, F. Rahmatpanah, S.H. Wei, and T.H. Huang. (2002) Applications of CpG island microarrays for high-throughput analysis of DNA methylation. *J. Nutr.* 132(8 Suppl), 2430S–2434S.
61. Hatada, I., M. Fukasawa, M. Kimura, S. Morita, K. Yamada, T. Yoshikawa, S. Yamanaka, C. Endo, A. Sakurada, M. Sato, T. Kondo, A. Horii, T. Ushijima, and H. Sasaki. (2006) Genome-wide profiling of promoter methylation in human. *Oncogene.* 25(21), 3059–3064.
62. Paulin R., G.W. Grigg, M.W. Davey, and A.A. Piper. (1998) Urea improves efficiency of bisulphite-mediated sequencing of 5'-methylcytosine in genomic DNA. *Nucleic Acids Res.* 26(21), 5009–5010.
63. Rein, T., H. Zorbas, and M. DePamphilis. (1997) Active mammalian replication origins are associated with a high-density cluster of mCpG dinucleotides. *Mol. Cell Biol.* 17, 416–426.
64. Clark, S.J., J. Harrison, C.L. Paul, and M. Frommer, (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.*, 22, 2990–2997.
65. Lipatov, A.S., E.A. Govorkova, R.J. Webby, H. Ozaki, M. Peiris, Y. Guan, L. Poon, and R.G. Webster. (2004) Influenza: Emergence and control. *J. Virol.* 78(17), 8951–8959.