Design of Tiling Arrays and Their Application to Bacterial Transcriptome Analysis

Takeshi Akama, Kazuaki Nakamura, Akito Tanoue, and Koichi Suzuki

Abstract

Whole-genome sequencing and annotation have clarified total gene number and structure in a variety of organisms. Microarrays have facilitated transcriptome analysis through the use of probes that target a large number of genes based on genomic information. However, microarrays are limited in that they can only examine known or predicted genes; non-annotated genes and noncoding regions cannot be accounted for.

Recent advances in technology have led to the design of tiling arrays, which contain a vastly increased number of spotted probes, and at higher density. Tiling arrays cover the entire genome of a prokaryotic species in an unbiased fashion by designing a large number of probes. Upon hybridization of total RNA, all the transcribed regions of the genome, irrespective of gene annotation, can be detected. As opposed to next-generation sequencing, tiling arrays are cost-effective, easy to analyze, and have been used for experiments as diverse as transcriptome analysis, ChIP-chip, and DNA sequence variation detection. In this chapter, the methods for bacterial tiling array slide design, RNA sample preparation, hybridization, and data analysis are described.

Key words Tiling array, Transcriptome, Bacteria, Genome, Noncoding region

1 Introduction

Microarray techniques permit the analysis of hundreds of thousands of genes in a single experiment by using probes mounted on a glass slide. Microarrays spotted with probes specific for every gene have been developed and applied as a high-throughput technique for analyzing gene expression in a variety of organisms [1–5]. However, only known genes are taken into account with probe design; other regions of the genome, including non-annotated coding regions, pseudogenes, and noncoding regions, cannot be analyzed in an open reading frame (ORF) array (Fig. 1). Noncoding regions, which do not encode functional proteins, were previously thought to be “junk” DNA and were not analyzed. However, there has been increased awareness of the functions of noncoding
regions, and some pseudogenes are known to be expressed and function in the regulation of gene expression [6–8].

In tiling arrays, probes are designed to span the entire genome of an organism without bias, regardless of whether or not the regions are annotated (Fig. 1). With typical eukaryotic genomes, the tiling array is designed for each genome contig because of the large genome size. Tiling arrays for prokaryotes, which have smaller sized genomes compared to eukaryotes, can be designed to cover the whole genome in an unbiased arrangement and thus can analyze the entire transcriptome. It has been applied to a variety of species such as *Bacillus subtilis* [9], *Neisseria meningitidis* [10], and *Saccharomyces cerevisiae* [11]. Previously, a whole-genome tiling array was designed for *Mycobacterium leprae* (*M. leprae*), a bacterium with a small genome and a large number of noncoding regions, and its transcriptome was analyzed [12]. This application of the tiling array for *M. leprae* is depicted as an example.

Fig. 1 Probe arrangement of the tiling array and ORF array. For the genome sequence of the target organism, probes for the tiling array are arranged at equal intervals, although the ORF array only probes known genes. Thus, not only the coding regions but also pseudogenes and other noncoding regions from the whole genome are covered by the tiled probes in a tiling array. Depending on the genome size and maximum probe number of the array slide, neighboring probes can be overlapping or separate.

### 2 Materials

#### 2.1 Tiling Array Design


2. 1× 4.2 M high density HD4 array (Roche NimbleGen, Madison, WI).

3. Custom array synthesis system (Roche NimbleGen).

#### 2.2 RNA Extraction from Bacteria

1. RNA Protect Bacteria Reagent (Qiagen, Germantown, MD).

2. 1.0 mm Zirconia Beads (BioSpec Products, Bartlesville, OK).
3. mirVana miRNA Isolation Kit (Ambion, Austin, TX).
5. DNase I (TaKaRa, Kyoto Japan).

2.3 Preparation of Labeled Double-Stranded DNA

1. SuperScript II (Life Technologies, Carlsbad, CA).
2. RNaseOUT (Life Technologies).
3. RNase A (Novagen, Madison, WI).
4. Cy3-9mer Wobble primer (TriLink Biotechnologies, San Diego, CA).

2.4 Hybridization and Analysis of the Tiling Array

2. MAUI Hybridization System (BioMicro Systems, Salt Lake City, UT).
4. Microarray High-Speed Centrifuge (Arrayit, Sunnyvale, CA).
5. GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA).
6. NIMBLESCAN 2.3 (NimbleGen Systems).

2.5 Quantitative Real-Time PCR

2. FastStart Universal SYBR Green Master (Rox) (Roche Diagnostics, Indianapolis, IN).
3. MicroAmp Optical 96-well Reaction Plate (Life Technologies).
4. MicroAmp Optical Adhesive Film (Life Technologies).
5. ABI Prism 7000 Sequence Detection System (Life Technologies).

3 Methods

The tiling array can identify unexpected transcriptional activity in a number of genomic regions, although the data obtained may have a lower signal-to-noise ratio. This is due to the fact that the probe sequences cannot be designed based on DNA specificity or GC content. Thus, proper probe length is important to improve the signal-to-noise ratio. The interval (degree of overlap) of the probes, which is decided by the genome size of the target organism and the number of probes mounted on the array slide, determines the resolution of expressed regions on the genome (Fig. 1). RNA degradation must be low during sample preparation as short segments of degraded RNA can hybridize nonspecifically.
1. The genome sequence of the organism to be analyzed can be obtained from the NCBI Web site, “Genomic DNA sequence.” The Web site GOLD can also be available to search both complete and incomplete genome projects. In case annotation is lacking, the genome sequence can be analyzed using the annotation programs on the GOLD Web site.

2. The genome sequence is divided into equal segments for probe design. The length of each probe is 60-mer and adjacent probes are spaced 18 nucleotides apart, resulting in probes that overlap every 42 nucleotides on the genome (Fig. 2) (see Note 1). Because the maximum allowable probe number is 4,200,000 in the Roche Nimblegen custom array, probe length and overlap between adjacent probes can be determined according to the specific genome length (see Note 2).

3. Random probes and control probes are also designed. Multiple random probes that have different sequences should be prepared to determine the background level. Control probes are hybridized with control nucleotides (Alignment Oligo of NimbleGen Hybridization Kit), spiked into samples, and used as an internal control to compare data obtained from different arrays.

4. Based on the above designs, the probes are synthesized on an array slide by Roche Nimblegen (see Note 3).

3.1 Tiling Array Design

1. Prepare bacterial cells yielding at least 1 μg RNA upon extraction. Suspend the cells in 2 ml of RNA Protect Bacteria Reagent and incubate for 10 min at room temperature. Centrifuge the cell suspension at 9,000×g for 6 min at 4 °C and remove the supernatant.

2. Resuspend the cell pellet in a mixture of 2 ml of RNA Protect Bacteria Reagent, 0.4 ml of 1.0 mm zirconia beads (see Note 4), and 0.6 ml of lysis/binding buffer from the mirVana miRNA Isolation Kit (see Note 5).

![Fig. 2 interval and overlap of the tiling array probe. For each of the two adjacent probes, 18 nucleotides of the 60-mer probe are shifted, resulting in an overlap of 42 nucleotides. Four consecutive probes share a partial sequence.](image-url)
3. Homogenize the mixture four times at 3,000 rpm for 3 min using a Micro Smash Homogenizer followed by a freeze–thaw cycle. Repeat four times.

4. Add 60 μl (0.1 volume of lysis/binding buffer) of homogenate additive to the homogenate and vortex. After a 10-min incubation on ice, add 600 μl (one volume) of acid–phenol:chloroform and vortex for 1 min. Centrifuge the sample at maximum speed for 5 min at room temperature to separate the organic and aqueous phases.

5. Preheat the elution solution to 95 °C at this time.

6. After centrifugation, transfer the supernatant to a new tube, add 750 μl (1.25 volumes) of ethanol, and mix at room temperature. Transfer the mixture (up to 700 μl at a time) to a filter cartridge that has been placed in a collection tube.

7. Centrifuge the cartridge at 10,000 × g for 15 s and discard the flow through. Repeat the centrifugation step in order to pass all the mixture through the cartridge.

8. Apply 700 μl of wash solution 1 to the filter cartridge and centrifuge at 10,000 × g for 10 s. Discard the flow through.

9. Apply 500 μl of wash solution 2/3 and centrifuge at 10,000 × g for 10 s. Discard the flow through. Repeat this step.

10. Centrifuge the cartridge at 10,000 × g for 10 s for 1 min and transfer the filter cartridge to a new collection tube.

11. Apply 100 μl of preheated (95 °C) elution solution to the center of the filter. Centrifuge the filter cartridge at maximum speed for 30 s.

12. Measure the absorbance of the eluent at 260 and 280 nm to determine RNA concentration and purity. The ratio of A$_{260}$/A$_{280}$ should be greater than 1.8.

13. Dispense one or more μg of RNA into a microtube and add 30 μl (0.1 volume) of DNase I buffer, 15 U of DNase I, and distilled water to a total volume of 300 μl. Incubate the mixture at 37 °C for 30 min followed by acid–phenol:chloroform extraction and ethanol precipitation. Dissolve the RNA pellet in 10 μl RNase-free water (see Note 6).

3.3 Preparation of Labeled Double-Stranded DNA

1. Mix up to 5 μg of total RNA with 250 ng random primer, 1 μl of 10 mM dNTP mix, and RNase-free water to a total volume of 12 μl. Incubate the mixture at 65 °C for 5 min and immediately transfer to ice.

2. Add 4 μl of 5× first-strand buffer, 2 μl of 0.1 M DTT, and 1 μl of RNaseOUT to the mixture. Incubate at 25 °C for 2 min.
3. Add 1 μl of SuperScript II RT to the mixture, and incubate at 25 °C for 10 min. Next, incubate the mixture at 42 °C for 50 min followed by heating at 70 °C for 15 min.

4. Incubate the synthesized cDNA from step 3 with 10 ng of RNase A at 37 °C for 10 min followed by phenol:chloroform extraction and ethanol precipitation.

5. Incubate 1 μg of ds-cDNA at 98 °C for 10 min with 1 OD\textsubscript{600} unit of Cy3-9mer Wobble primer (see Note 7) and distilled water to a total volume of 80 μl.

6. Add 10 μl of dNTPs (10 mM each) and 2 μl (100 U) of Klenow fragment and incubate at 37 °C for 2 h. Stop the reaction by adding 10 μl (0.1 volume) of 0.5 M EDTA, and precipitate the labeled cDNA with isopropanol.

3.4 Hybridization of the Tiling Array

1. Warm the MAUI Hybridization System to 42 °C for 3 h prior to use.

2. Resuspend the Cy3-labeled sample in 1.7 μl of distilled water plus 3.3 μl of sample tracking control, 9.0 μl of 2× NimbleGen Hybridization Buffer, 3.6 μl Hybridization Component A, and 0.37 μl of Alignment Oligo.

3. Denature the mixture by heating at 95 °C for 5 min and maintain at 42 °C until sample loading.

4. Remove an X1 Mixer from its package 30 min prior to use. Using the Precision Mixer Alignment Tool (PMAT), place the tiling array slide on the mixer.

5. Place the mixer-slide assembly in the slide bay of the MAUI Hybridization System. Apply the sample through the fill port until it starts to leak from the vent port. Remove excess sample that has leaked from the mixer and adhere a mixer port seal over the fill and the vent ports.

6. Hybridize the samples for 18 h at 42 °C using mix mode B of the MAUI Hybridization System.

7. Prepare wash I, II, and III by mixing 225 ml of distilled water, 25 μl of 1 M DTT, and 10× wash buffer I, II, or III.

8. Load the mixer-slide assembly into the Mixer Disassembly Tool immersed in wash I at 42 °C. After peeling off the mixer, agitate the slide for 10–15 s.

9. Transfer the slide to a processing container containing wash I. Shake the processing container at least 20 times every 10 s. After 2 min, transfer the slide into wash II and agitate for 1 min. Next, transfer the slide into wash III and agitate for 15 s.

10. Remove the slide from the wash III and dry for 1 min using the microarray high-speed centrifugation. Prior to scanning, keep the slide shielded from light.
3.5 Scanning of the Array Slide

1. Turn on the GenePix 4000B scanner and launch the GenePix software 10 min prior to use.

2. Blow compressed gas across the slide to remove any contaminants. Place the slide in the slide carriage so that the array is face down and the barcode end is closest to you.

3. Move the black lever on the left side of the carriage to the left until the slide is lying flat in the carriage. Release the lever so that the slide is gently pushed to the right side of the carriage and held firmly in place.

4. Close the slide carriage and shut the scanner door.

5. On the Hardware Settings dialog box of GenePix software, select the following settings: Wavelength to 532, PMT gain to 500, Power to 100, Pixel size to 5, Lines to average to one, and Focus position to zero.

6. Click the Scan Area icon and draw a box that surrounds the array image by dragging the mouse. The box should be slightly larger than the array image.

7. Click the play icon to scan the array slide.

8. Adjust the PMT gain so that normalized counts at the 65,000-intensity level are 1e-5. Stop the scan and restart under the new setting.

9. Save the obtained image. Remove the array slide from the scanner and store in a dark desiccator for rescanning.

10. Launch the NimbleScan software. The scanned image, a design file (.ndf), and a gene description file (.ngd) will open.

11. Perform the auto brightness/contrast adjust function and the auto align tool.

12. Adjust the grid so that the corner fiducial controls line up correctly with the grid.

13. Run the local alignment tool to fine-tune the alignment and save the file.

14. Create a pair file by selecting Reports-Pair in the Analysis menu, which contains the raw data and lists the probe intensities of the array. The gridded image file is assigned the same .ndf file as the one in step 10.

15. Create a general feature format file (.gff) by selecting Reports-GFF in the Analysis menu, which can be opened with the SignalMap software. A position file (.pac) will be assigned (see Note 8).

3.6 Data Analysis on SignalMap Software

1. Launch the SignalMap software and the tiling array data file will open with a gene annotation file.

2. The positive regions of hybridized signals can be detected by applying the Find Peaks function in the Track menu. Shorter
positive regions can be detected by decreasing the Peak window size, although some positive regions may be fragmented. Percent of Peak Threshold affects the sensitivity. Positive regions that have lower signal intensity can be detected by decreasing the Percent of Peak Threshold (Fig. 3).

3. In the line created after applying the Find Peaks function, positive regions are shown as rectangles. The number of Rows indicates the number of positive regions. The precise range and highest score of the positive region are shown when hovering the mouse over the rectangles.

**3.7 Quantitative Real-Time PCR**

1. The cDNA used for the tiling array can also be subjected to real-time PCR analysis.

2. The primers can be designed using Primer-BLAST, Primer3, or other primer design programs. Primers are designed to amplify 80–300 bp PCR products due to the amplification efficiency of target genes.

3. To generate mean and standard error values of the relative cDNA amounts, prepare the reactions in triplicate. For one of the reactions, mix the cDNA with 10 μl of 2× FastStart Universal SYBR Green Master (Rox), 0.04 μl of 100 μM forward and reverse primer, and distilled water to a total volume of 20 μl. Furthermore, prepare the reactions in duplicate using a series of diluted sample for the standard cDNA to draw standard curve.
Apply the mixtures to a MicroAmp Optical 96-well Reaction Plate and seal the top of the plate with MicroAmp Optical Adhesive Film.

Turn on ABI PRISM 7000, the computer, and launch the ABI PRISM 7000 SDS software.

Create a new document for Absolute Quantification and set detectors which are selected SYBR in the reporter dye menu for each of the PCR products in the Detector Manager window.

Define the Sample name, Detector, Task, and quantity options in the Well Inspector dialog box on the Plate tab. The same sample names are filled for triplicate samples. For the standard curve, mixtures of a series of diluted sample are selected in the Task option to STD and dilution factors are filled in the Quantity option.

Under the Instrument tab, set the sample volume to 20 μl and check the box labeled Dissociation Protocol. Use the default protocol setting on the Thermal Cycler: an initial step of 2 min at 50 °C and 10 min at 95 °C followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C.

After saving the file, place an Optical Cover Compression Pad on the top of the plate and set the plate to ABI PRISM 7000. Close the lid of the machine and start the run.

When the run is finished, check the Standard Curve and Dissociation curve. To certify that there are no PCR by-products, the Dissociation Curve should have single peak at the same position among triplicates. After exporting the results, the relative amounts of cDNA may be obtained. Examples of the comparison between the quantitative real-time PCR result and the tiling array result are shown in Fig. 4.

**Fig. 4** Comparison of RNA expression between real-time PCR and the tiling array. For positive regions in genes, pseudogenes, and other noncoding regions, relative RNA expression levels detected by quantitative real-time PCR (left vertical axis) are compared with the signal intensities detected by the tiling array (right vertical axis). All of the positive regions in the tiling array can be detected by real-time PCR; however, the magnitude of their expression levels differs with the result from the tiling array.
4 Notes

1. Longer probes have higher specificity, but lower resolution. Although 29-mer probes are spotted for some predesigned ORF arrays, they give vaguer signal patterns than 60-mer probes. By aligning the tiling array data and ORF array data, it is shown that specific signals can be obtained by the tiling array with 60-mer probes like ORF array as shown in Fig. 5.

2. Probes can be designed to both the plus and minus strand, as long as the total probe number is under half of the maximum of the custom array slide. In this design, probe specificity can be checked with the data from both strands.

3. Tiling Array Tools (http://www.affymetrix.com/partners_programs/programs/developer/TilingArrayTools/index.affx) (Affymetrix, Santa Clara, CA) can also be used.

4. For some bacterial species, lysozyme treatment may be used instead of homogenization. Resuspend the cell pellet in 2 ml of RNA Protect Bacteria Reagent, add 0.4 ml of 1.2 mg/ml lysozyme in TE, and incubate for 5 min at room temperature. Centrifuge the mixture at 5,000×g for 10 min. Repeat the lysozyme incubation step once more, and add 0.6 ml lysis/binding buffer from the mirVana miRNA Isolation Kit. The remaining steps are the same from Subheading 3.2, step 4.

5. The mirVana miRNA Isolation Kit is used to extract total RNA including short RNA. Other RNA extraction reagents like TRIZOL (Life Technologies, Carlsbad, CA) can also be used.

6. If an excess amount of RNA has been extracted, checking for RNA degradation by electrophoresis is recommended. Prepare samples for loading by mixing several μg of total RNA with 2.5 μl of 10× MOPS (MERCK, Darmstadt, Germany), 1 μl of 1.0 mg/ml ethidium bromide (Sigma), 4.5 μl formaldehyde, and 12.5 μl formamide. Incubate the mixture at 65 °C for 15 min and chill on ice. Add 2.5 μl of RNA gel loading solution (Quality Biological, Inc., Gaithersburg, MD), and load the samples on a 1 % agarose gel with 1× MOPS running buffer. An optimal ratio of 23S/16S rRNA amount should be greater than 1.0.

7. Cy3-labeled random nonamer may be used instead of Cy3-9mer Wobble primer. Dissolve 1 O.D. of primer in 42 μl of 12.5 mM Tris–HCl, pH7.4, 1.25 mM MgCl₂, and 25 mM β-mercaptoethanol.

8. The values of signal intensity are supplied as raw data in the .gff file. To convert them to log2 values, open the .gff file in Microsoft Excel and use the LOG function. The modified file can be saved as a tab-delimited text file and can be changed from .txt to .gff.
Fig. 5 Influence of probe length on signal specificity. For *M. leprae* genomic DNA (GENBANK Acc. No. NC_002677), the tiling arrays spotted with 60- or 29-mer probes were designed. The obtained signal intensity is graphed on the vertical axis and each probe is horizontally arrayed against the corresponding genome sequence using the SignalMap software. The data from a 60-mer array are shown in the upper position, and the data from a 29-mer array are shown in the middle position. White rectangles positioned underneath indicate the range of coding regions and gray rectangles indicate the range of pseudogenes. Broad (a) and narrow (b) ranges of the genome are shown. (c) The data from the 60-mer tiling array are aligned with the data from the ORF array. In the ORF array, different 20 probes are designed against one coding region to ensure sequence specificity.
References

Tiling Arrays
Methods and Protocols
Lee, T.-L.; Shui Luk, A.C. (Eds.)
2013, XI, 236 p. 53 illus., 37 illus. in color., Hardcover
ISBN: 978-1-62703-606-1
A product of Humana Press