Improved Quantitation and Reproducibility in *Mycobacterium tuberculosis* DNA Microarrays

Benjamin G. Schroeder*1, Liane M. Peterson, and Robert D. Fleischmann

The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850
1Current address: Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404

Abstract

We show here that the amount of labeled cDNA and its specific activity can play a significant role in the quantitation of microarray experiments. Standard reverse transcription of 2 μg total bacterial RNA with concomitant incorporation of cyanine dye-conjugated nucleotides did not produce enough label for optimal hybridization results in our *Mycobacterium tuberculosis* DNA microarray. Therefore we turned to an alternative labeling method using the incorporation of aminolyl nucleotides followed by conjugation to Cy-dye. The method allows up to 10 fold more label to be produced, and at higher specific activity. In particular, more transcripts can be detected and variability between replicate features can be reduced by using more labeled cDNA. We show that optimizing the labeling protocol is a critical element in conducting microarray experiments and obtaining reproducible and interpretable data.

Communication

The use of spotted DNA microarrays to measure relative gene transcription has surged in recent years; however, little has been published on factors that affect the outcome of such experiments. Before meaningful quantitative data can be produced from any experimental system, it is important to have at least some basic understanding of how different aspects of the experimental protocol affect the data produced. In the course of developing a microarray of *Mycobacterium tuberculosis*, we sought to understand the relationship between the amount of probe used and the quality of the data generated from the array. [In this paper, we use the term probe to refer to the material labeled (RNA reverse transcribed into cDNA), and target to refer to the amplicons spotted on the array.]

Our microarray consists of purified PCR amplicons from each open reading frame predicted from the genome sequence of the recent clinical isolate, *M. tuberculosis* strain CDC1551 (Valway et al., 1998, http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.jsp?database=gmt). Six open reading frames present in strain H37Rv (Cole et al., 1998) but absent in CDC1551 were also included. A total of 4067 gene specific primers were used to amplify a portion of each open reading frame, ranging in size from 26 to 753 bp in length. The quality of each amplicon was examined first by agarose gel electrophoresis, and ultimately by dye-terminator sequencing using the same primers used to generate the amplicon. PCR products were purified in MANU30 96-well filter plates (Millipore), resuspended in water, mixed 1:1 with DMSO and spotted onto aminosilane-coated high contact angle glass slides (Corning) in duplicate with ChipMaker 3 pens (TeleChem International, Sunnyvale, CA) using a robot from Intelligent Automation Systems (Cambridge, MA). After spotting, microarrays were cross-linked with 250 mJ of UV radiation in a Stratalinker (Strategene) and stored at room temperature in a vacuum dessicator.

We found that the amount of probe used has a large effect on the sensitivity and reproducibility of a microarray experiment (Figure 1). We prepared large amounts of probe from two RNA samples using the aminolyl labeling protocol described below and a 2:3 molar ratio of aa-dUTP:dTTP. Five microarrays were hybridized with increasing amounts of these probes. Probes were hybridized in 50% formamide, 5 × SSC, 0.2% SDS, 1 mg/ml salmon testes DNA, and 12.5 mM EDTA in a final volume of 40 μl under a 25 mm × 60 mm LifterSlip (Erie Scientific) at 42°C for 18 hours. Prehybridization and washing were performed as described in Hegde et al. (2000), with the addition of 10 mM EDTA to the prehybridization solution. We measured sensitivity by counting the number of genes for which hybridization signals were detected. Features were quantitated with TIGR Spotfinder (available at http://www.tigr.org/softlab/). In order to be counted as expressed, both duplicate features were required to have intensity in both channels. As shown in Figure 1, the number of expressed genes rose from 742, when 20 pmol each of Cy3 and Cy5 were used, to 2080 when 140 pmol of each dye were used. The number of expressed genes that could be detected did not increase with the addition of more probe, suggesting all gene transcripts present in the RNA sample had been identified.

We also used the duplicate features on each array to measure the variability, expressed as percent error. To calculate the percent error, we divided the absolute difference of the Cy3/Cy5 ratios of two features by the average ratio of the same two features. Figure 1 shows the median of the percent error plotted versus the amount of probe. The data show that increasing
Further Reading

Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at caister.com

- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof

- **Aspergillus and Penicillium in the Post-genomic Era**
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2

- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorfl, SM Roy, MA Rile (2016)
  www.caister.com/bacteriocins

- **Omics in Plant Disease Resistance**
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr

- **Acidophiles: Life in Extremely Acidic Environments**
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles

- **Climate Change and Microbial Ecology: Current Research and Future Trends**
  Edited by: J Marxsen (2016)
  www.caister.com/climate

- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
  Edited by: G Lear (2016)
  www.caister.com/bioremed

- **Microalgae: Current Research and Applications**
  Edited by: MN Tsagolou (2016)
  www.caister.com/microalgae

- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
  www.caister.com/gasplasma

- **Virus Evolution: Current Research and Future Directions**
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusvol

- **Arboviruses: Molecular Biology, Evolution and Control**
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo

- **Shigella: Molecular and Cellular Biology**
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella

- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
  Edited by: AM Romani, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms

- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alphha

- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile

- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow

- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmon (2015)
  www.caister.com/probiotics

- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics

- **Corynebacterium glutamicum: From Systems Biology to Biotechnological Applications**
  Edited by: A Burkovski (2015)
  www.caister.com/cory2

- **Advanced Vaccine Research Methods for the Decade of Vaccines**
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines

- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals

- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
  www.caister.com/bacteria-plant

- **Aeromonas**
  Edited by: J Graf (2015)
  www.caister.com/aeromonas

- **Antibiotics: Current Innovations and Future Trends**
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics

- **Leishmania: Current Biology and Control**
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2

- **Acanthamoeba: Biology and Pathogenesis (2nd edition)**
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2

- **Microarrays: Current Technology, Innovations and Applications**
  Edited by: Z He (2014)
  www.caister.com/microarrays2

- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order
the amount of probe decreases the percent error of duplicate features. We conclude that increasing the amount of labeled probe in a microarray hybridization experiment yields two benefits. First, it increases the likelihood that all transcripts present in the RNA sample will be observed, and second, it reduces the variability in the measured expression ratios.

It has been our experience that the presence of Cy-dCTP or Cy-dUTP inhibits reverse transcription by SuperScript II (Invitrogen) reverse transcriptase. Typical reactions using random primers and Cy-dUTP (Peterson et al., 2000) would incorporate 10–20 pmol of Cy dye into cDNA per microgram of total RNA template, with correspondingly low percent yields of 10 to 30 percent. It became clear from the experiments described above that in order to generate the highest quality microarray data we needed to generate much more probe than the standard Cy-nucleotide protocol would afford. The aminoallyl labeling protocol (using 5-(3-aminoallyl)-2'-deoxyuridine 5'-triphosphate, aa-dUTP) proved to be an attractive alternative. The method was developed from two online protocols, http://cmgm.stanford.edu/pbrown/protocols/aadUTP-CouplingProcedure.htm, and http://www.pangloss.com/seidel/Protocols/amino-allylRT.html. In our standard aminoallyl labeling protocol, 2 µg of total RNA are mixed with 6 µg of random hexamers and denatured at 70°C for 10 minutes, then snap-cooled in a dry ice/ethanol bath. Reverse transcription was performed at 42°C for 3 hours in a total of 30 µl containing 0.5 mM each dATP, dCTP and dGTP, 400 units SuperScript II RNase H⁻ reverse transcriptase, 10 mM DTT and 1× first strand buffer (Invitrogen). Various ratios of dTTP and aa-dUTP (Sigma) were used, but the combined concentration was always 0.5 mM. After reverse transcription, RNA was hydrolyzed at 65°C for 15 minutes by the addition of 10 µl each 1 M NaOH and 0.5 M EDTA. The cDNA was neutralized with the addition of 25 µl 1 M Tris-HCl, pH 7.2. Unincorporated aa-dUTP and free amines were removed by using a modified QIAquick PCR spin column (QIAGEN) purification. The purification was similar to the manufacturer’s protocol, except that the wash buffer was replaced with 5 mM

<p>| Table 1. Results of cDNA labeling with various ratios of aminoallyl-dUTP:dTTP. |
|-----------------|-------|--------|--------|</p>
<table>
<thead>
<tr>
<th>aa-dUTP:dTTP</th>
<th>% Yield</th>
<th>pmol Cy3</th>
<th>nt/dye</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:4</td>
<td>165</td>
<td>250</td>
<td>41</td>
</tr>
<tr>
<td>2:3</td>
<td>130</td>
<td>464</td>
<td>17</td>
</tr>
<tr>
<td>1:1</td>
<td>174</td>
<td>685</td>
<td>16</td>
</tr>
<tr>
<td>3:2</td>
<td>156</td>
<td>769</td>
<td>13</td>
</tr>
<tr>
<td>4:1</td>
<td>144</td>
<td>762</td>
<td>12</td>
</tr>
<tr>
<td>5:0</td>
<td>90.6</td>
<td>749</td>
<td>7.5</td>
</tr>
</tbody>
</table>

aResults shown are from one experiment. The experiment was repeated with a different RNA sample with similar results.
bMolar ratio. [aa-dUTP] + [dTTP] was kept at 0.5 mM.
cCalculated as ng cDNA synthesized divided by ng RNA template used.
dNucleotides per Cy3.
potassium phosphate, 80% ethanol, pH 8.0, and the elution buffer was replaced with 4 mM potassium phosphate, pH 8.5. Eluted cDNA was dried in a vacuum concentrator, resuspended in 4.5 μl 0.1 M sodium carbonate, pH 9.0 and mixed with 4.5 μl of FluoroLink Cy5 or Cy3 monofunctional dye (Amersham Pharmacia Biotech) dissolved in DMSO.

Each FluoroLink dye pack was resuspended in 72 μl of DMSO and used immediately, or aliquoted and stored at −80°C. After 1 hour at 25°C, 35 μl of 0.1 M sodium acetate, pH 5.2 were added to the reaction, and the Cy-labeled cDNA was purified with a QIAquick PCR spin column according to the manufacturer’s instructions.

Figure 2. Frequency polygon of ratios of cDNA Cy3 feature intensity to Cy5 genomic DNA control standard for each feature. For each microarray, the same mole amount of fluorophore was used and the nucleotide per dye ratio was changed. Diamonds, squares, and triangles correspond to high, medium, and low nucleotide per dye ratios, respectively, of the cDNA probes. Panels A and B show the results from two experiments using two different RNA samples. The leftmost datapoint depicts the number of features with Cy3/Cy5 ratios below 0.01. The next datapoint depicts the number of features with ratios between 0.01 and 0.02, and so on.
With this labeling method, percent yields are significantly increased (see Table 1), suggesting that aa-dUTP is less inhibitory to reverse transcriptase than Cy-dUTP. The molar ratio of aa-dUTP to dTTP can be altered to achieve the desired specific activity. Table 1 shows how altering this ratio affects the percent yield, specific activity, and total dye incorporation. Increasing the ratio of aa-dUTP to dTTP results in increased total dye incorporation and increased specific activity (lower nucleotide per dye value). The reaction even works when all dTTP is replaced with aa-dUTP, resulting in very specific activity probe (7.5 nt/dye in this experiment).

Next, we investigated how the specific activity of a probe affects signal intensity. In order to measure and compare signal intensities across different probes and different array hybridizations, a genomic DNA standard labeled with Cy5 was hybridized along with each Cy3-labeled cDNA to each array. Genomic DNA was isolated from M. tuberculosis strain CDC1551 using the bead beatng protocol (Jacobs et al., 1991) and labeled by the incorporation of Cy3- or Cy5-dCTP by random priming with exoKlenow using a protocol loosely based on that of Behr, et al. (1999). Five μg of genomic DNA and 15 μg of random hexamers were denatured at 95°C for 5 minutes then snap-cooled in a dry ice/ethanol bath. Labeling was performed at 25°C for 3 hours in a total of 50 μl, containing 0.12 mM each dATP, dGTP and dTTP, 0.06 mM dCTP, 0.06 mM Cy5-dCTP (Amersham Pharmacia Biotech), 25 units exoKlenow and 1×EcoPol buffer (New England Biolabs). Labeled DNA was purified with a QiAquick PCR spin column. As an aside, hybridization with labeled genomic DNA is an excellent way to measure the number of hybridization-competent features on an array. Experiments of this kind demonstrate that over 95% of the features on our M. tuberculosis microarray are hybridization competent (data not shown).

Since the same Cy5 standard was used in all hybridizations, the different Cy3-labeled cDNA probes could be compared by measuring the Cy3 to Cy5 ratio. Two experiments were performed using different samples of M. tuberculosis RNA for each. In both cases, three different specific activities were compared, ranging from 41 nucleotides per Cy3 to 7.2 nucleotides per Cy3. Figure 2 shows the frequency distributions of feature intensities for the three different specific activity cDNA probes, expressed as the ratio of the cDNA Cy3 intensity to genomic DNA Cy5 intensity. The distributions were very consistent between the two experiments. In both cases the cDNA probes with 40 and 20 nucleotides per dye gave similar intensity distributions, while the distribution of 7 nt/dye probes was shifted to lower values, indicating these probes gave less intense signals.

An important benefit of using larger amounts of probe is the increased reliability of expression data (Figure 1). To achieve a similar level of labeling using Cy-dCTP protocols would require scaling up the reaction 5- to 10-fold. We calculate the Cy-dCTP labeling reaction costs about $18, while a reaction using the aa-dUTP protocol described here costs about $14. Scaling up to provide probes comparable to the aa-dUTP method results in considerable expense and is an inefficient use of bacterial RNA samples.

We also show that the aminoallyl labeling protocol allows a wide range of aa-dUTP concentrations to be used, resulting in a wide range of specific activity probes. This labeling strategy also has the advantage that it can be used to couple any label that is available as a NHS ester. This allows the use of other fluorophores besides Cy3 and Cy5, including labels that are not currently available as nucleotide conjugates or are inhibitory to reverse transcriptase.

With the ability to generate very high specific activities, we investigated the effect of specific activity on microarray hybridizations. We found that at the highest specific activity (7 nt/dye) the intensity of fluorescence dropped off (Figure 2). One possible explanation for this is that the high density of fluorophores on the cDNA probe interferes with hybridization. Another possibility is that the high density of fluorophores leads to self-quenching of fluorescence. For M. tuberculosis, which has a G+C content of 65.6%, we routinely use a 2:3 mixture of aa-dUTP:dTTP. This provides a good balance between the competing desires for keeping the nucleotide per dye ratio in the 20 to 40 range, generating as much probe as possible, and using as little of the expensive aa-dUTP as possible (see Table 1). By performing the experiments described here it should be possible to easily identify the optimal labeling conditions for RNA from other species with different G+C content.

Acknowledgements

We thank Deepak Kaushal and William Bishai for M. tuberculosis RNA, Vasily Sharov, Alexander Saeed, Joseph White, Jianwei Li and John Quackenbush for software and database help, and Jeremy Peterson for primer design software.

References


