

when more than two groups were involved. If data were not normally distributed, a nonparametric test (Mann–Whitney U-test) was used for the comparisons of results. Data were expressed as mean  $\pm$  s.e.m.

### Acknowledgments

The R1 ES cell line was kindly provided by Dr. Heiner Westphal. The second ES cell line used E14.1 was a gift of Dr. Tom Doetschman. We would like to thank Dr. David Panchinson for providing some of the PCR-primers and for critical discussions of the manuscript.

- Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- Reubinoff, B. et al. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat. Biotechnol.* **18**, 399–404 (2000).
- Shamblott, M.J. et al. Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc. Natl. Acad. Sci. USA* **95**, 13726–13731 (1998).
- Klug, M.G. et al., Genetically selected cardiomyocytes from differentiating embryonic stem cells form stable intracardiac grafts. *J. Clin. Invest.* **98**, 216–24 (1996).
- Keller, G. et al. Overexpression of HOX11 leads to the immortalization of embryonic precursors with both primitive and definitive hematopoietic potential. *Blood* **92**, 877–87 (1998).
- Wobus, A.M. & Boheler, K.R. Embryonic stem cells as developmental model in vitro. *Cells Tiss. Org.* **165**, 129–30 (1999).
- Bain, G. et al. Embryonic stem cells express neuronal properties in vitro. *Dev. Biol.* **168**, 342–357 (1995).
- Deacon, T. et al. Blastula-stage stem cells can differentiate into dopaminergic and serotonergic neurons after transplantation. *Exp. Neurol.* **149**, 28–41 (1998).
- Brustle, O. et al. Embryonic stem cell-derived glial precursors: a source of myelinating transplants. *Science* **285**, 754–756 (1999).
- Okabe, S. et al. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mech. Dev.* **59**, 89–102 (1996).
- Olanow, C.W., Kordower, J.H. & Freeman, T.B. Fetal nigral transplantation as a therapy for Parkinson's disease. *Trends Neurosci.* **19**, 102–109 (1996).
- Breier, A., et al., Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: evidence from a novel positron emission tomography method. *Proc. Natl. Acad. Sci. USA* **94**, 2569–2574 (1997).
- Gramm, L.F. Drug therapy: fluoxetine. *N. Engl. J. Med.* **331**, 1354 (1994).
- Volkow, N.D. et al. Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects. *Nature* **386**, 830–833 (1997).
- Volkow, N.D. et al. Relationship between subjective effects of cocaine and dopamine transporter occupancy. *Nature* **386**, 827–830 (1997).
- Simeone, A. Otx1 and Otx2 in the development and evolution of the mammalian brain. *EMBO J.* **17**, 6790–6798 (1998).
- Acampora, D. & Simeone, A. Understanding the roles of Otx1 and Otx2 in the control of brain morphogenesis. *Trends Neurosci.* **22**, 116–122 (1999).
- Stoykova, A. & Gruss, P. Roles of Pax genes in developing and adult brain as suggested by expression patterns. *J. Neurosci.* **14**, 1395–1412 (1994).
- Rowitch, D.H. & McMahon, A. P., Pax-2 expression in the murine neural plate precedes and encompasses the expression domains of Wnt-1 and En-1. *Mech. Dev.* **52**, 3–8 (1995).
- Lendahl, U., Zimmermann, L.B. & McKay, R.D.G. CNS stem cells express a new class of intermediate filament proteins. *Cell* **60**, 585–595 (1990).
- Frederiksen, K. & McKay, R.D.G. Proliferation and differentiation of rat neuroepithelial precursor cells in vivo. *J. Neurosci.* **8**, 1144–1151 (1988).
- Johe, K. et al. Single factors direct the differentiation of stem cells from the fetal and adult nervous system. *Genes Dev.* **10**, 3129–3140 (1996).
- Vicario-Abejón, C. et al. Functions of basic-fibroblast growth factor and neurotrophins in the differentiation of hippocampal neurons. *Neuron* **15**, 105–114 (1995).
- Studer, L., Tabar, V. & McKay, R. Transplantation of expanded mesencephalic precursors leads to recovery in parkinsonian rats. *Nat. Neurosci.* **1**, 290–295 (1998).
- Studer, L., Tabar, V. & McKay, R.D.G. Survival of expanded dopaminergic precursors is critical for clinical trials. *Nat. Neurosci.* **1**, 537 (1998).
- Ye, W. et al. FGF and SHH signals control dopaminergic and serotonergic cell fate in the anterior neural plate. *Cell* **93**, 755–766 (1998).
- Branton, R.L., Love, R.M. & Clarke, D.J. cAMP included during cell suspension preparation improves survival of dopaminergic neurons in vitro. *Neuroreport* **9**, 3223–3227 (1998).
- Kalir, H.H. & Mytilineou, C. Ascorbic acid in mesencephalic cultures: effects on dopaminergic neuron development. *J. Neurochem.* **57**, 458–464 (1991).
- Brustle, O. et al. In vitro-generated neural precursors participate in mammalian brain development. *Proc. Natl. Acad. Sci. USA* **94**, 14809–14814 (1997).
- Spenger, C. et al. Fetal ventral mesencephalon of human and rat origin maintained in vitro and transplanted to 6-hydroxydopamine-lesioned rats gives rise to grafts rich in dopaminergic neurons. *Exp. Brain Res.* **112**, 47–57 (1996).
- Wagner, J. et al. Induction of a midbrain dopaminergic phenotype in Nurr1-overexpressing neural stem cells by type 1 astrocytes. *Nat. Biotechnol.* **17**, 653–659 (1999).
- Joyner, A.L. Engrailed, Wnt and Pax genes regulate midbrain–hindbrain development. *Trends Genet.* **12**, 15–20 (1996).
- Studer, L. et al. Noninvasive dopamine determination by reversed phase HPLC in the medium of free-floating roller tube cultures of rat fetal ventral mesencephalon: a tool to assess dopaminergic tissue prior to grafting. *Brain Res. Bull.* **41**, 143–150 (1996).
- Auerbach, J.M., Eiden, M.V. & McKay, R.D.G. Transplanted CNS stem cells form functional synapses in vivo. *Eur. J. Neurosci.* **12**, 1696–1704 (2000).

# Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis

Adel M. Talaat, Preston Hunter, and Stephen Albert Johnston\*

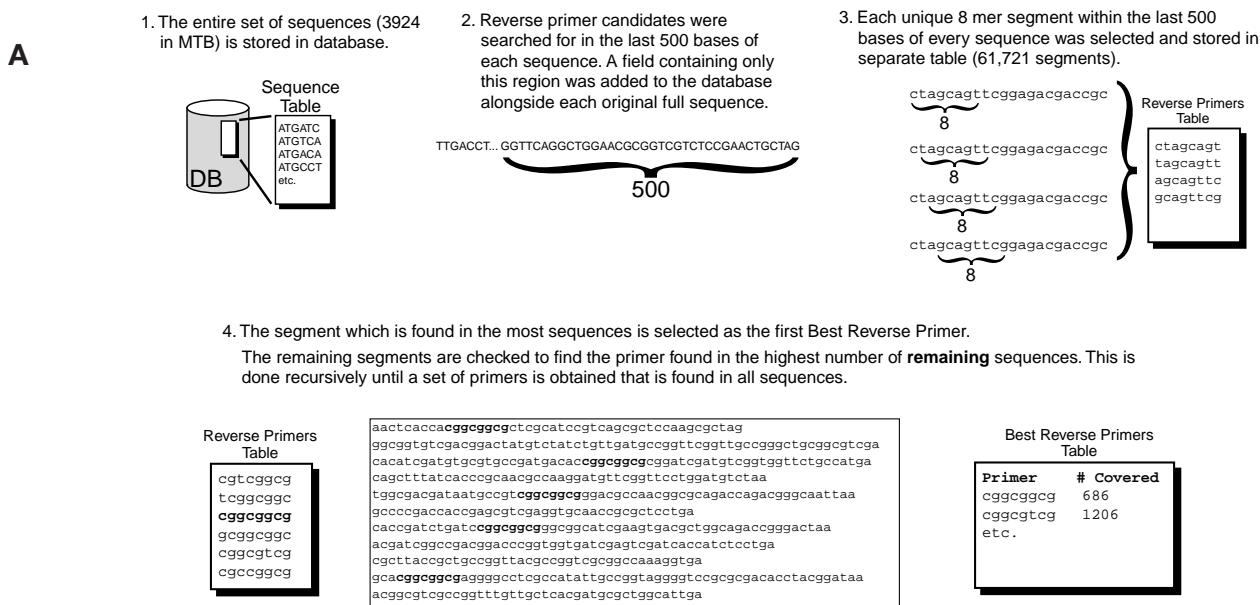
Center for Biomedical Inventions and Department of Medicine, University of Texas-Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas TX 75390-8573. \*Corresponding author ([johnston@ryburn.swmed.edu](mailto:johnston@ryburn.swmed.edu))

Received 9 March, 2000; accepted 25 April, 2000

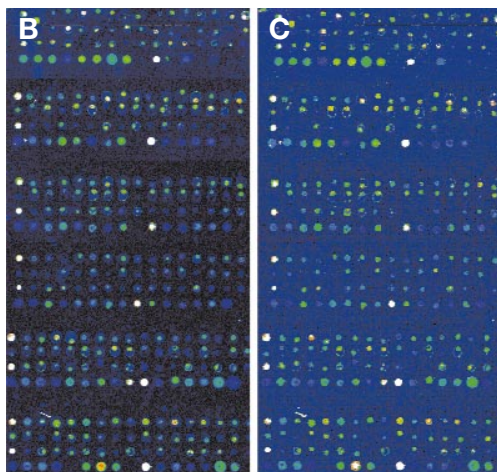
DNA microarrays have the ability to analyze the expression of thousands of the same set of genes under at least two different experimental conditions<sup>1</sup>. However, DNA microarrays require substantial amounts of RNA to generate the probes, especially when bacterial RNA is used for hybridization (50  $\mu$ g of bacterial total RNA contains approximately 2  $\mu$ g of mRNA)<sup>2</sup>. We have developed a computer-based algorithm for prediction of the minimal number of primers to specifically anneal to all genes in a given genome. The algorithm predicts, for example, that 37 oligonucleotides should prime all genes in the *Mycobacterium tuberculosis* genome. We tested the usefulness of the genome-directed primers (GDPs) in comparison to random primers for gene expression profiling using DNA microarrays. Both types of primers were used to generate fluorescent-labeled probes and to hybridize to an array of 960 mycobacterial genes. Compared to random-primer probes, the GDP probes were more sensitive and more specific, especially when mammalian RNA samples were spiked with mycobacterial RNA. The GDPs were used for gene expression profiling of mycobacterial cultures grown to early log or stationary growth phases. This approach could be useful for accurate genome-wide expression analysis, especially for *in vivo* gene expression profiling, as well as directed amplification of sequenced genomes.

We developed a computer algorithm to define the minimal number of oligonucleotides of a given length capable of priming all genes within any genome. Using the genome sequence of *Mycobacterium tuberculosis*, we applied the algorithm, setting the oligonucleotides length at eight or seven bases, and requiring 100% coverage of the 3,924 open reading frames (ORF) in the genome<sup>3</sup> (Fig. 1A). The search was limited to the first 500 bp of each complementary sequence of each ORF to generate long probes for efficient hybridization. The priming efficiency of the mycobacterial genome-directed primers (mtGDPs; 37 primers) was compared to the priming efficiency of seven- or six-nucleotide random primers in a standard reverse transcription reaction. Probes generated from the same mycobacterial RNA (log phase cultures) using mtGDPs or random primers were hybridized simultaneously to the same slides, and the signal intensities of the 960 arrayed genes were calculated and compared<sup>4,5</sup>. There was a high correlation level between the signals of both probes for the whole array of genes ( $r = 0.97$ ). Additionally, signals generated by mtGDPs were significantly higher than random primer-generated signals ( $P < 0.05$ ) (Fig. 1B, C). Signal intensities

## TECHNICAL REPORTS



**Figure 1 (A)** Flowchart depicting the algorithm used for developing genome-directed primers from bacterial genomes. *Mycobacterium tuberculosis* genome was used as a model for such analysis. **(B, C)** Comparison of cDNA labeling using different protocols. A fluorescence image of a subset of 960-gene array hybridized with labeled cDNA primed with either mtGDPs **(B)** or random primers **(C)**. Each image represents six constellations of arrayed genes, where the first spot in rows 1 and 3 is a purified 16S rDNA-PCR product.



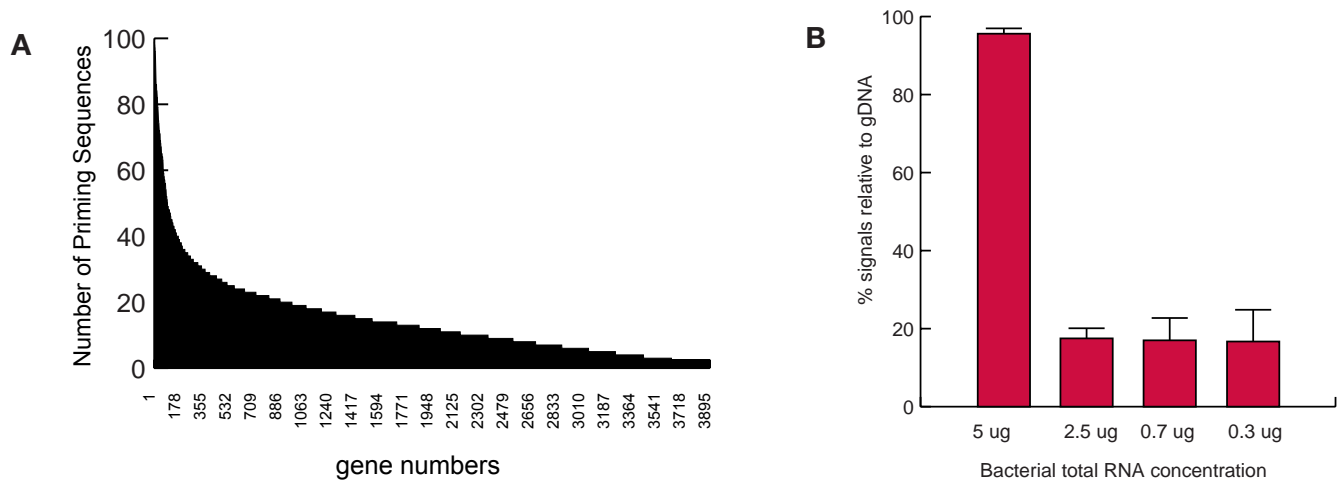
To test the usefulness of the mtGDPs for *in vivo* gene expression analysis, we asked whether the mtGDPs can be used for labeling bacterial RNA when mixed with mammalian RNA (extracted from activated monocytic cell lines<sup>6,7</sup>). In these experiments, Cy5-labeled probes were generated from 5 µg of pure mycobacterial RNA whereas Cy3-labeled probes were generated from 5 µg of mycobacterial RNA mixed with mammalian RNA in different ratios (1:50 and 1:100). Genes with signal intensities generated from spiked mammalian RNA probes that did not have corresponding true signals from the pure mycobacterial probes were considered genes with nonspecific signals (possibly generated from the overlapping sequences in mammalian RNA). Using the mtGDPs probes at 60°C hybridization, nonspecific signals account-

ed for 25% and 27% of the true signals when 50:1 and 100:1 spiked mammalian RNA were used, respectively (data not shown). However, at 65°C hybridization, the nonspecific signals dropped to 13.5% of the true signals. In contrast, using the same stringent hybridization conditions (65°C) and probes generated from random primers, nonspecific signals accounted for 32% of the true signals when 50:1 spiked mammalian RNA was used.

A possible reason for genes with nonspecific signals observed with mixed RNA samples is the sequence similarity between mycobacterial transcripts and mammalian transcripts from activated monocytic cells. Upon mycobacterial infection and especially in the site of infection, the host RNA will be rich with cytokine transcripts<sup>2</sup>. Therefore, we analyzed the frequency of occurrence of mtGDPs in human cytokine-encoding genes. Based on sequence identity, our computer analysis revealed the majority of mtGDPs have a 0–5% chance of priming cytokine transcripts. To directly test this overlap, we arrayed 34 human cytokine encoding genes alongside mycobacterial genes on the same slide. Equal amounts (5 µg) of mammalian RNA extracted from activated THP-1 and U-937 cells were labeled in standard reverse transcription reactions using either random primers or mtGDPs. The labeled cDNA probes were mixed and allowed to hybridize to the same DNA microarrays at 60°C. High signal intensities were observed for the mammalian genes when probed with random primer-generated cDNA whereas signals were below the background levels for the mtGDP-generated cDNA

from both probes were above the threshold level of the background (true signals) in 86% (mtGDPs) to 89% (random primers) of the arrayed genes. Random primers with seven or six nucleotides gave similar results when compared to the mtGDPs.

To test for any biased priming of mtGDPs toward any class of genes, we computed the sequence distribution of the mtGDPs set in the *M. tuberculosis* genome. Based on this sequence homology (Fig. 2A), 82 % of the genes could be primed with the mtGDPs at 10 or more sites, whereas only 2.5% of the genes could be primed at 100 or more sites. However, after probe hybridization, we found no correlation between the computed priming matches of each gene and the signal intensities of such genes. For example, the 16S rRNA gene with 10 hits of mtGDPs gave the highest expression level under all growth conditions, whereas other genes with more than 100 hits of mtGDPs gave low expression levels. To decide on a suitable RNA concentration for labeling using the GDPs, decreasing concentrations of mycobacterial total RNA (5.0, 2.5, 0.7, and 0.3 µg) were used to generate probes for DNA microarrays. The total number of DNA spots with true signal intensities generated from either RNA-based probes or genomic DNA probes were compared after hybridization to the 960 DNA microarrays. Less than 20% of the genes generated true signals from 0.2–2.5 µg of total RNA compared to 96% when 5 µg total RNA were used to synthesize the labeled probe (Fig. 2B). Furthermore, an overall reduction in signal intensity for each gene accompanied the reduction in the amounts of mycobacterial RNA.



**Figure 2. mtGDPs for measuring gene expressions. (A)** A histogram representing the distribution of mtGDPs in the *M. tuberculosis* genome. The numbers of mtGDPs match in each gene were sorted from the highest to the lowest. Genes with more than 100 matches of mtGDPs per gene were excluded (0.5% of the genes). **(B)** A histogram representation (with standard errors) of true hybridization signals generated from 0.3, 0.7, 2.5, or 5.0  $\mu\text{g}$  of total *M. tuberculosis* RNA in comparison to hybridization signals generated from 2  $\mu\text{g}$  *M. tuberculosis* genomic DNA.

(data not shown). This demonstrates the relative lack of priming of at least this set of mammalian genes by mtGDPs under moderately stringent hybridization conditions.

We next used the mtGDPs to identify differentially expressed genes (*degs*) in mycobacterial cultures grown to logarithmic phase compared to stationary phase. For this set of experiments, Cy5-labeled genomic DNA (gDNA) from *M. tuberculosis* H37Rv was cohybridized with Cy3-labeled cDNA prepared from either logarithmic or stationary-phase RNA. Signals generated from gDNA allowed the normalization of gene expression level since every gene is present once in the genome. Normalized ratios of log phase probes and stationary-phase probes were used to calculate the log ratios of gene expression under different growth conditions. Genes with a change in log ratio of threefold or more from the mean are considered genes with high expression levels (Table 1 and our web site<sup>8</sup>). About 10% of genes were differentially expressed under both growth conditions. The majority of the *degs* were of unknown function.

**Table 1. Partial list of genes differentially expressed in *M. tuberculosis* growing in log or stationary phases<sup>a</sup>**

Stationary phase		Log phase	
Gene name	Log ratio	Gene name	Log ratio
rpmJ	0.74	MTV005.08c	0.87
rpIP	0.64	MTV025.093c	0.80
IS1547	0.58	MTV025.001B	0.69
rpIK	0.43	MTV027.15c	0.66
ribF	0.37	MTV034.04	0.65
cysA	0.36	MTYC373.10c	0.59
nrdG	0.35	MTV004.03c	0.47
galU	0.35	MTV025.026c	0.43
cysE	0.35	<i>RecX</i>	0.37
rpsS	0.32	MTV014.24c	0.36
<i>pstB</i>	0.32	<i>Eph</i>	0.35
hspR	0.31	MTV017.53	0.33
aroQ	0.31	<i>DapA</i>	0.32
clpP2	0.30	MTV021.08c	0.31

<sup>a</sup>The normalized ratios of intensity for Cy3- (total RNA) to Cy5 (genomic DNA)-labeled probes were used as a measure for transcripts hybridization signals for each gene relative to its gDNA hybridization signals. A logarithmic ratio of the normalized simple ratios was calculated to compare the expression profile of mycobacterial cultures grown at log and stationary phases. A change of threefold or more from the mean of the normalized fluorescence intensities of each probe was considered a significant change of gene expression. The complete table can be accessed from our web site<sup>8</sup>. <http://cbi.swmed.edu/Computation/gdp/gdp.html>.

In bacterial gene expression studies, random primers are usually used to prepare reverse transcriptase-dependent cDNA. However, when using arrays for estimating gene expression levels during the pathogen infection, only limited amounts of bacterial mRNA contaminated with the host RNA are present. Under such circumstances, random primers would label both the host and pathogen transcripts, thereby decreasing the sensitivity of detecting bacterial mRNA and increasing the chances for nonspecific signals. In this report, we investigated the usefulness of a strategy that uses a minimal set of oligonucleotides deduced from a sequenced genome to prime all transcripts for bacterial DNA microarrays analysis. The rationale was that such genome-directed primer sets would preferentially prime the pathogen transcripts, leading to the detection of the pathogen transcripts in a heterogeneous population of RNA. Labeled cDNA prepared from mtGDPs was more efficient and more sensitive in assays of mycobacterial transcripts, even when mycobacterial RNA was diluted with mammalian RNA. In mixed RNA samples and under stringent hybridization conditions, a lower percentage of nonspecific signals (13.5%) were detected when mtGDPs were used to generate the labeled probes as compared to random primers (32%). If necessary, the assay could be improved by changing the hybridization conditions or altering the algorithm to eliminate primers recognizing mammalian transcripts. However, more stringent hybridization conditions could reduce the overall hybridization signals, making the signal quantification of the whole array more difficult. By eliminating primers responsible for crosshybridization, a higher number of primers would be needed. We also demonstrated that mtGDPs could be used to identify differentially expressed genes of *M. tuberculosis* in stationary versus log growth phases.

The GDP algorithm could also be used for finding primers for PCR amplification for rapid building of genomic libraries. Using the mtGDPs, it should be possible to use RNA samples extracted from infected tissues or cell cultures in DNA microarray assays without any further purification, enrichment, or concentration procedures. Our results indicate that GDPs may provide more sensitivity and specificity in the analysis of pathogen transcripts in infected tissues, facilitating searches for new targets for vaccine and drug therapy.

### Experimental protocol

**Design of mycobacterial genome-directed primers (mtGDPs).** The mtGDPs are 7- or 8-mer oligonucleotides chosen from *M. tuberculosis* genome using a novel searching algorithm (Fig. 1A). The algorithm can be used to define the minimal number of oligonucleotides of a given length capable of priming all

## TECHNICAL REPORTS

genes within any genome. The basic algorithm parameters are (1) sequences of each gene in a genome (2) the length of the search region (in either forward or reverse directions) within which an oligonucleotide must have a match (3) the desired number and length of the oligonucleotides, and (4) the percentage of the ORFs to be covered. We applied this algorithm to *M. tuberculosis* genome using the sequence data stored in an SQL database. The 37 oligonucleotides identified through this method as well as the GDP-Finder program can be downloaded from our web site<sup>8</sup>. These primers were synthesized in-house using ABI DNA Synthesizer (Applied Biosystems, Foster City, CA) and used for labeling mycobacterial transcripts from purified mycobacterial total RNA or from mixed mycobacterial and mammalian total RNA in different ratios as described below.

**Hybridization and data processing.** A detailed description of building the mycobacterial DNA microarrays and generating labeled probes can be found in our web site<sup>8</sup>. For each hybridization experiment, cDNA-labeled probes were generated from 5 µg of total RNA harvested from mycobacterial cultures grown to either log or stationary growth phases. Labeling of RNA transcripts was performed in presence of reverse transcriptase (SuperScript II, Gibco/BRL, Grand Island, NY), 6-mers random primers (Gibco/BRL), 7-mers random primers (ABI DNA Synthesizer), and GDPs (ABI DNA Synthesizer) were labeled with Cy3- or Cy5-dCTP fluorophores (Amersham, Arlington Heights, IL) according to the manufacturer's protocol. When decreasing concentrations of mycobacterial RNA (less than 5 µg) were tested, total RNA extracted from Cos-7 cells was used to maintain 5 µg for each labeling reaction. Mycobacterial genomic DNA samples (2 µg each) were labeled by nick translation system labeling kit (Promega, Madison, WI) and according to the manufacturer's protocol. Hybridized slides were scanned by a laser confocal microscopy (Scan Array 3000, Telechem International Inc., San Jose, CA) after independent excitation of the fluorophores Cy3 and Cy5. The signal and background fluorescence intensities were calculated for each DNA spot using image analysis software (Imagene, Biodiscovery Inc., Los

Angelas, CA) by averaging the intensities of every pixel inside the target region (segmentation method). The probe intensity for each DNA spot was the difference between average signal intensity and average local background intensity. The ratios of intensity for Cy5- to Cy3-labeled probes were determined for each DNA spot, reflecting the relative abundance of mRNA transcripts in each probe. A change of two standard deviations or more from the mean of the normalized fluorescence intensities of each fluorophore was considered a significant preferential gene expression.

### Acknowledgments

*We wish to thank Walker Hale for excellent technical support, Chuck Epstein and Ross Chambers for helpful discussions. We particularly thank Susan Howard and Rick Lyons for providing RNA and DNA samples from M. tuberculosis H37Rv cultures. This work was supported by grants from DARPA and NIH to S.A.J. A.M.T. is supported in part by a cardiology training grant fellowship.*

1. Carulli, J.P. et al. High throughput analysis of differential gene expression. *J. Cell. Biochem.* 286–296 (1998).
2. De Saizieu, A. et al. Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays *Nat. Biotechnol.* 16, 45–48 (1998).
3. Cole, S.T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–538 (1998).
4. Schena, M. Genome analysis with gene expression microarrays. *Bioassays* 18, 427–431 (1996).
5. Eisen, M.B. & Brown, P.O. DNA arrays for analysis of gene expression. *Methods Enzymol.* 303, 179–205 (1999).
6. Eperon, S. & Jungi, T.W. The use of human monocytoid lines as indicators of endotoxin. *J. Immunol. Methods* 194, 121–129 (1996).
7. Yao, J., Mackman, N., Edgington, T.S. & Fan, S.T. Lipopolysaccharide induction of the tumor necrosis factor-alpha promoter in human monocytic cells: regulation by Egr-1, c-Jun, and NF-kappaB transcription factors. *J. Biol. Chem.* 272, 17795–17801 (1997).
8. <http://cbi.swmed.edu/Computation/gdp/gdp.html>