

## Optimizing RNA extraction yield from whole blood for microarray gene expression analysis

Jian Wang, John F. Robinson, Hafiz M.R. Khan, David E. Carter, James McKinney, Brooke A. Miskie, and Robert A. Hegele\*

*Vascular Biology Research Group and London Regional Genomics Centre, Robarts Research Institute and University of Western Ontario, London, ON, Canada N6A 5K8*

Received 3 March 2004; accepted 25 March 2004

Available online 27 April 2004

### Abstract

**Objectives:** Microarray analysis of gene expression profiles of blood leukocytes has many potential clinical and research applications.

**Design and methods:** We used the PAXgene Blood RNA System to prepare RNA from the whole blood of normal volunteers using two incubation times followed by gene expression profiling using the Affymetrix HU133A GeneChip.

**Conclusions:** Longer incubation gave a significantly higher RNA yield and samples that were satisfactory for microarray analysis, with excellent pairwise correlations between replicates.

© 2004 The Canadian Society of Clinical Chemists. All rights reserved.

*Keywords:* Molecular diagnosis; Gene expression; Immunology; Epidemiology; Metabolic disease

### Introduction

Leukocytes from whole blood are routinely used for research and diagnostics in hematological, metabolic, or genetic diseases [1]. Profiling of gene expression using microarrays is a new analytical platform for biomedical data, especially from accessible tissues such as blood leukocytes. However, the peculiarities of RNA as a biological material, especially degradability, may compromise traditional blood sample collection and preparation methods [2]. To begin to understand some technical issues surrounding analysis of RNA expression profiles in peripheral blood leukocytes, we undertook a pilot study to determine RNA yield and gene expression profiles from whole blood. The PAXgene Blood RNA Tubes and Blood RNA Kit system (QIAGEN, Mississauga, ON) was selected because the protocol allowed for standardized sample collection directly into vacuum tubes containing lysis buffer and a stabilization solution, which is a mixture of cationic detergent and salts. This method was previously shown to yield RNA of suf-

ficient quantity and quality for reverse transcriptase-based analysis of one or two genes at a time [3]. Based on that report, we evaluated the suitability of this method for preparation of RNA from whole blood and analyzed the prepared RNA using the expression profiling of the whole genome on the Affymetrix Inc. (Santa Clara, CA) platform.

### Methods

#### *Study subjects*

We obtained multiple whole blood samples from healthy volunteers after informed consent as approved by the Human Ethics Board, University of Western Ontario. Baseline blood samples were taken after a 12-h fasting period and samples were collected on a second day after a 12-h fasting period for biological replicate comparisons. There was no change in health, diet, or medication for study subjects over this time period.

#### *Blood collection and RNA preparation*

Blood was collected directly into PAXgene collection tubes using a 21-gauge butterfly needle and catheter so

\* Corresponding author. Blackburn Cardiovascular Genetics Laboratory, Robarts Research Institute, Room 406, 100 Perth Drive, London, ON, Canada N6A 5K8. Fax: +1-519-663-3037.

E-mail address: [hegele@robarts.ca](mailto:hegele@robarts.ca) (R.A. Hegele).

that tubes could be held low and vertical to maintain the vacuum on top of the stabilization solution. This allowed the blood to flow directly into the solution with no time delay. Tubes were inverted 10 times. Each tube drew a maximum of 2.5 ml of blood, and between four and six tubes were taken at each time point. Samples were incubated in collection tubes at room temperature for either 4 or 24 h, which were the extremes of the time range specified in the manufacturer's instructions. Total RNA was isolated according to the manufacturer's instructions. The entire procedure was carried out at room temperature with the exception of proteinase K incubation at 50°C for 10 min. RNA was eluted from the PAXgene column according to the manufacturer's instructions. The total RNA yield was determined by absorbance at 260 nm.

#### Microarray gene expression profiles

Total RNA (10 µg) was used for labelling and then 15 µg of labelled cRNA was hybridized to each Human Genome U133A array (Affymetrix Inc.) according to the manufacturer's instructions and protocols of the London Regional Genomics Centre ([www.lrgc.ca](http://www.lrgc.ca)). The "subject 1, day 1" sample was a pool of the RNA extracted from one subject after a 4-h incubation, while the "subject 1, day 2" sample was a pool of the RNA collected on a different day from the same subject and extracted after a 24-h incubation. In contrast, samples "subject 2, day 1" and "subject 2, day 2" were each a pool of RNA collected on respective days from a second subject that had each been extracted after a 24-h incubation. Expression profiles were analyzed using D-Chip perfect match-only model [4] utilizing the outlier and artefact characterization features. We then used the filtering algorithms in three different microarray analysis software programs, namely Affymetrix Microarray Suite (MAS) version 5.1 Data Mining Tool (DMT) v3.0, GeneSpring version 5.0.2 (Silicon Genetics, Redwood City, CA), and Bullfrog [5]. These filtering algorithms allowed us to generate lists of genes that showed 1.5-fold differences up or down between paired samples from the same subject. For each pairwise comparison, the log transformations of the signal ratios were correlated against each other.

## Results

#### Quantity and quality of prepared RNA

Eighteen RNA samples were studied with respect to RNA yield per ml blood collected according to incubation time in PAXgene tubes at room temperature: either 4 h ( $N = 5$ ) or 24 h ( $N = 13$ ). Mean total RNA yields per tube for the 4- and 24-h incubation were significantly different: respectively,  $2.53 \pm 0.52$  and  $4.33 \pm 0.68$  µg/ml blood,  $P < 0.0001$  (Fig. 1A).

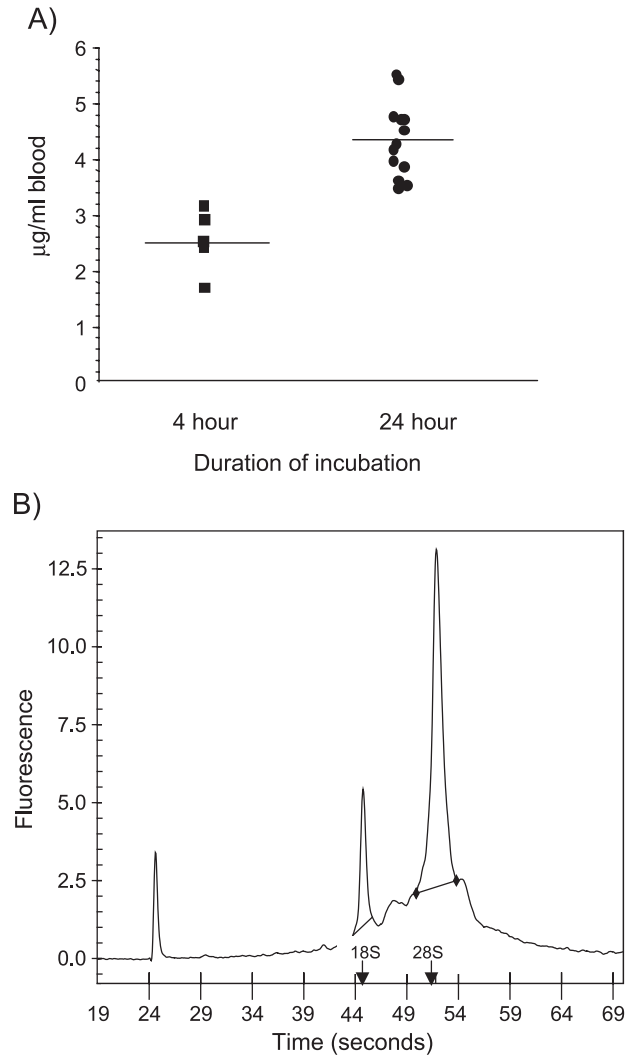


Fig. 1. (A) RNA yield according to incubation period. Individual sample yields are shown with solid symbols and means are indicated with lateral lines. Mean total RNA yields per tube for the 4- and 24-h incubation were significantly different: respectively,  $2.53 \pm 0.52$  ( $N = 5$ ) and  $4.33 \pm 0.68$  ( $N = 13$ ) µg/ml blood,  $P < 0.0001$ . (B) Total RNA electropherogram profile from Agilent Bioanalyzer. A representative profile of total RNA extracted from a sample that had been incubated for 24 h is shown. Positions of 18S and 28S rRNA are shown with arrows. The leftmost peak is the marker. No extraneous peaks are seen in this tracing.

RNA extractions were precipitated to achieve a final RNA concentration of  $>20$  µg/10 µl. RNA integrity was confirmed using the Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA) individually before and after precipitation according to the manufacturer's instructions. When compared with electropherograms of cRNA from other tissues, cRNA obtained from whole blood showed two peaks instead of one, including a peak of about 700 base pairs (data not shown) that has been described previously [1]. Electropherograms of total RNA from all samples had a uniform profile consistent with adequate RNA quality for expression analysis, with no extraneous or unexpected peaks (Fig. 1B).

### Gene expression profiles

RNA extractions from two subjects from two different days (biological replicates) went on to microarray analysis. Four hybridizations were performed in total (“subject 1, day 1”, “subject 1, day 2”, “subject 2, day 1”, and “subject 2, day 2”). D-Chip analysis of expression profiles showed that the per-chip outlier values were all below the warning limits of 5%, indicating no assay-induced artefacts. For replicate samples from the same subject, pairwise analysis was performed to characterize the data distribution. Sample “subject 1, day 1” was compared to sample “subject 1, day 2”, with the former serving as the normalization sample. Sample “subject 2, day 1” was compared to sample “subject 2, day 2”, with the former serving as the normalization sample. For all four samples, the percent of total 22,283 possible gene expression values registered as “present” ranged from 48.2 to 51.2, which was comparable to results reported for other RNA isolation and extraction protocols [1,2,6].

Using the filtering algorithms in Affymetrix DMT v3.0, GeneSpring v5.0.2, and Bullfrog [5], we generated lists of genes that showed 1.5-fold differences up or down between paired samples. For each pairwise comparison, the log transformations of the signal ratios were plotted, resulting in a normal distribution (mean ratio  $0.05 \pm 0.26$ ), showing that the vast majority of the changes in gene expression were below the  $\pm 2\sigma$  limit. The very small numbers of up- and downregulated genes (Table 1) increased our confidence that the samples represented true biological replicates. These percentages were even smaller when stricter criteria, such as a 2-fold

changes for pairwise comparisons, were used (data not shown). GeneSpring analysis tended to identify a higher percentage of genes with more extreme changes than 1.5-fold for both pairwise comparisons. In addition, all three programs indicated a markedly higher percentage of 1.5-fold downregulated genes for comparisons involving the sample that had been incubated for only 4 h.

### Discussion

We describe a standardized method to extract and prepare excellent quality total RNA from whole blood using specific conditions within the PAXgene Blood RNA Tube and RNA Kit systems platform. This application of the PAXgene reagents produced sufficient total RNA for Affymetrix GeneChip hybridization that resulted in data sets with acceptable percent present calls for gene expression profiles. We found that a 24-h incubation of cells in PAXgene tubes at room temperature produced a higher RNA yield than previously described using incubation times in PAXgene tubes between 2 and 9 h [2,6]. Furthermore, the 24-h incubation of cells produced a significantly higher RNA yield—about 50% higher—than the 4-h incubation. The molecular or biochemical basis for such increased RNA yield is not clear, but might be a topic to explore in future studies. In contrast to the results from others [6], we did not observe additional peaks either attributable to DNA or otherwise unaccounted for on the bioanalyzer tracings, which were uniform for all RNA samples. After labelling, hybridization, and gene expression profiling, our observed percent “present” calls (approximately 50%) were within the range seen in our facility using RNA extracted from a variety of human tissue sources (range 45–60% based on the first 311 microarray experiments performed in our facility). While there was a very good pairwise comparison of expression profiles between replicate samples from the same subject under fasting conditions on a different day, there appeared to be somewhat fewer 1.5-fold downregulated calls in comparison between samples that had each been incubated for 24 h. The significance of this finding is not known at present and might require verification using a larger number of study samples and a more systematic approach.

Therefore, we report an application of a minimally invasive method that permits analysis of gene expression profiles from whole blood. For smaller volume samples, such as those from children, it may be worth considering the potential benefit associated with the apparent increase in yield related to using an incubation time at the extreme end of the recommended range in the PAXgene Blood RNA Tube and RNA Kit systems protocol. We conclude that this procedure will be useful for future research or diagnostic applications that require RNA expression profiling.

Table 1  
Percentages of 22,283 transcripts that had greater than 1.5-fold differences between replicate samples

Sample	Microarray analysis software					
	Affymetrix DMT		GeneSpring		Bullfrog	
	% up	% down	% up	% down	% up	% down
Subject 1, day 1 (4 h) vs. subject 1, day 2 (24 h)	0.58	3.03	2.27	8.52	0.94	3.45
Subject 2, day 1 (24 h) vs. subject 2, day 2 (24 h)	0.16	0.20	2.19	2.76	0.27	0.39

Abbreviations: times in brackets indicate individual incubation times for the biological replicate samples from two subjects, as described in Methods. % up and % down refer to the proportion of the total 22,283 transcripts that differed by more than, respectively, 1.5-fold up and down, between the biological replicates. This limit is commonly used in screening for significant expression changes between two conditions (e.g., treated and untreated) in microarray experiments. The percent that did not change by more than 1.5-fold was therefore 100%–% up or 100%–% down. For instance, for a value of 0.58% up, 99.42% of transcripts did not vary by more than 1.5-fold between biological replicates.

## Acknowledgments

Dr. Hegele holds a Canada Research Chair (Tier I) in Human Genetics and a Career Investigator award from the Heart and Stroke Foundation of Ontario. This work was supported by grants from Canadian Institutes for Health Research (MT13430), the Canadian Genetic Diseases Network, the Ontario Research and Development Challenge Fund (the Facility for Advanced Genetic Analysis), and the Blackburn Group.

## References

- [1] Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, et al. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci U S A* 2003; 100:1896–901.
- [2] An analysis of blood processing methods to prepare samples for GeneChip® expression profiling. *Affymetrix Newsletter*, Technical note, July 2003.
- [3] Rainen L, Oelmueller U, Jurgensen S, Wyrich R, Ballas C, Schram J, et al. Stabilization of mRNA expression in whole blood samples. *Clin Chem* 2002;48:1883–90.
- [4] Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 2001;98:31–6.
- [5] Zapala MA, Lockhart DJ, Pankratz DG, Garcia AJ, Barlow C, Lockhart DJ. Software and methods for oligonucleotide and cDNA array data analysis. *Genome Biol* 2002;3(6) [SOFTWARE0001, Epub 2002 May 23].
- [6] Thach DC, Lin B, Walter E, Kruzelock R, Rowley RK, Tibbetts C, et al. Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. *J Immunol Methods* 2003;283:269–79.