

Assessment of normal variability in peripheral blood gene expression¹

Catherine Campbell, Suzanne D. Vernon*, Kevin L. Karem, Rosane Nisenbaum and Elizabeth R. Unger
Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

Abstract. Peripheral blood is representative of many systemic processes and is an ideal sample for expression profiling of diseases that have no known or accessible lesion. Peripheral blood is a complex mixture of cell types and some differences in peripheral blood gene expression may reflect the timing of sample collection rather than an underlying disease process. For this reason, it is important to assess study design factors that may cause variability in gene expression not related to what is being analyzed. Variation in the gene expression of circulating peripheral blood mononuclear cells (PBMCs) from three healthy volunteers sampled three times one day each week for one month was examined for 1,176 genes printed on filter arrays. Less than 1% of the genes showed any variation in expression that was related to the time of collection, and none of the changes were noted in more than one individual. These results suggest that observed variation was due to experimental variability.

1. Introduction

Arrays allow for simultaneous detection of hundreds to thousands of expressed genes [3]. Genome-wide analysis of gene expression increases our understanding of the varied relationships between biologic pathways and disease processes. To date, gene expression studies have focused on samples derived from cells or tissues of a known lesion. However, not all diseases have a defined lesion amenable to sampling (e.g. occult infections or diseases of the central nervous system). In disease states with no known lesions, the peripheral blood is as an ideal sample to profile since it is a complex mixture of cells whose expression reflects ongoing systemic processes. However, some of the peripheral blood gene expression could be due to factors independent of specific disease states, such as circadian

and hormonal rhythms, sleep, stress, mood and diet [1, 5,6,11]. For this reason, it is important to define the extent of normal variability in gene expression of the peripheral blood.

This study was designed to test whether there were changes in gene expression related to the time of peripheral blood collection. To address this, samples of circulating peripheral blood mononuclear cells (PBMCs) were collected from healthy volunteers three times one day each week for one month. The expression of 1,176 genes printed on high-density filter arrays was evaluated.

2. Methods and materials

2.1. Sample collection and preparation

Three medication-free volunteers with no known health problems were recruited from the laboratory (one female and two males, ages 29–60) and gave informed consent. Subjects were instructed not to smoke, eat, or drink for at least three hours prior to sampling. Venipuncture into citrate tubes was performed at 6 a.m., noon, and 6 p.m. one day a week for one month and the blood was processed within 2 hours of collec-

¹Human experimentation guidelines of the U.S. Department of Health and Human Services were followed in the conduct of this study. All study participants were volunteers who gave informed consent.

*Address for correspondence: Suzanne D. Vernon, Ph.D., Centers for Disease Control and Prevention, Mail Stop A-15, 1600 Clifton Road, Atlanta, GA 30333, Tel.: +1 404 639 1338; Fax: +1 404 639 2779; E-mail: svernon@cdc.gov.

tion. PBMCs were separated from other blood components using Lymphocyte Separation Media (LSM, ICN Biomedical, Aurora OH). The 7–8 mls of whole blood was gently diluted to 15 ml with physiologic saline, layered over 15 ml of LSM and centrifuged at 1500 RCF for 15 minutes at 23°C. The plasma was removed and the mononuclear cell band above the LSM collected and washed with physiological saline (Dulbecco's Phosphate buffered saline (DPBS), Invitrogen, Carlsbad, CA). The washed PBMCs were counted, and frozen for viability in aliquots of 5×10^6 cells in 1 mL complete media (RPMI, with 2 mM L-glutamine, 10 mM HEPES, (Invitrogen) and 10% fetal bovine serum, (Hyclone, Logan, UT)) with 20% DMSO (Sigma, St. Louis MO). Samples were stored in liquid nitrogen until ready for extraction.

2.2. RNA extraction

Two aliquots of PBMCs were thawed and rinsed once with complete media, twice with DPBS and pelleted at $800 \times g$ for 5 minutes. Cell counts and viability with trypan blue were assessed by hemocytometer. All aliquots had greater than 90% viability, and yielded more than 3.5×10^6 cells. Thawed cells were lysed with a QiaShredder Column (Qiagen, Valencia, CA) and RNA was extracted using the RNA Easy Kit (Qiagen). RNA quality and quantity were determined by UV spectrophotometry and denaturing agarose gel electrophoresis. The ratio of 28S to 18S rRNA bands in the stained ethidium bromide gel was determined using the Alpha Innotech Image Station and FluorChem 2.0 software (Alpha Innotech, San Leandro, CA). Any sample with a 28:18S ratio less than 1.2 was not used for further analysis.

2.3. Labeling and hybridization

A double stranded digoxigenin labeled cDNA probe was synthesized from 1 μg of total RNA as described previously [10]. Half of the labeled PCR product (50 μL) was used to hybridize to 1,176 genes printed on the Atlas Human 1.2II cDNA Expression Arrays (CLONTECH, Palo Alto, CA). Hybridization and chemiluminescent detection of the arrays was done as previously described [7,10] with the addition of a 2 minute, 90°C pre-wash of the filters in 0.5% SDS prior to prehybridization. Chemiluminescent signal was detected using a one-hour exposure in the FluorChem Image Station (Alpha Innotech). Exposure time was selected to maximize the number of detectable genes without significantly increasing background. Raw images (no contrast or brightness adjustment) were saved in TIFF format for analysis

2.4. Analysis

The TIFF image was analyzed with GenePix Pro 3.0 (Axon, Union City CA). For image analysis, a grid was made for each filter to localize and quantitate raw intensity, which is measured as the mean intensity of the total spot area minus local background. The raw intensity values for each array were imported into Excel (Microsoft) and scaled to a percent value using the lowest negative and highest positive controls on that array. A COV was determined for each array by averaging the intensity of 10 negative features plus 3 standard deviations. These 10 features were selected to represent background variation across the membrane and were the same features for each array including 4 negative control genes and 1 unexpressed gene from each of the 6 quadrants. Intensities below the COV were set to 0.01.

Three hundred and two genes that were negative at all time points for all three subjects were excluded from further analysis. To correct for variation between filters, the mean intensity of the remaining 874 genes was calculated for each filter. The average of the mean intensity for all filters was determined. Each filter was normalized by multiplying each feature's intensity by the ratio of the average of all filter means to each individual filter's mean. The normalized intensity values were transformed by \log_2 for analysis.

Variability in gene expression was examined over the course of a day and over a month for each subject. Subjects were not compared to each other. To assess variability over a day, pairwise comparisons of the four-week average for each time point were used: 6 a.m. to noon, 6 a.m. to 6 p.m., and noon to 6 p.m. For weekly comparisons, the intensity values for the three time points in a day were averaged and similarly used in pairwise comparisons: week 1 to 2, 1 to 3, 1 to 4, 2 to 3, 2 to 4, and 3 to 4. The overall agreement for each comparison was evaluated using the Lin concordance coefficient [4] a coefficient that measures the extent of variation from the line of identity. Greater than four-fold variation in individual gene intensity was considered significant. The variation in the expression of each gene was further examined with a paired t-test, corrected for multiple comparisons using the step-down method described by Westfall and Young [12]. Statistical analysis was performed using Microsoft Excel and SAS (SAS Institute, Cary, NC).

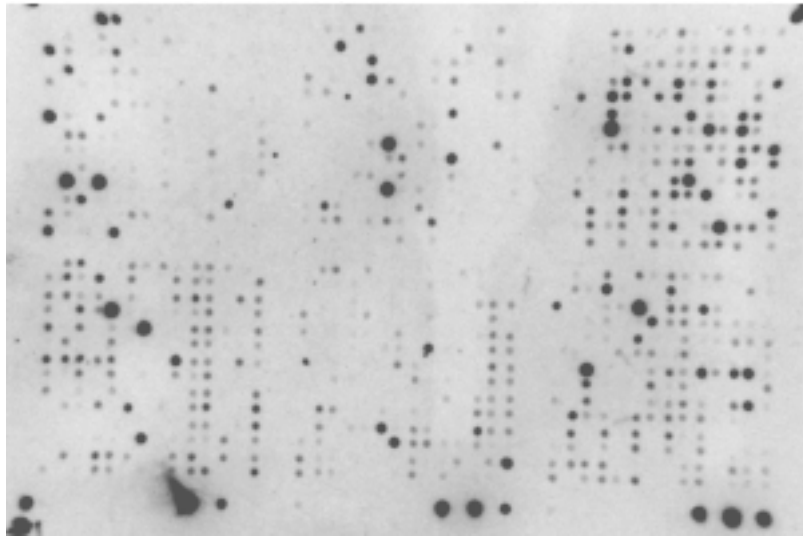


Fig. 1. A representative image of the 1.2II arrays obtained from the Alpha Innotech Image Station using FluorChem 2.0 software.

3. Results

3.1. Variability in gene expression by time of day

A typical result of hybridization and chemiluminescent detection on a filter array is illustrated in Fig. 1. Each individual's gene expression over the course of a day was found to have minimal variation in the expression of the majority of genes (concordance coefficient 0.98–0.99). Less than 0.5% (4/874) varied more than 4-fold over the 12 hour period within any one individual (Fig. 2 A–C).

Expression of only two genes varied significantly within the day (paired t-test, $p < 0.01$ with correction for multiple comparisons). This observation was not consistent for all subjects since one individual showed no significant variation and two different genes were identified in each of the other two individuals. In addition, neither of these genes showed a four-fold difference in expression.

3.2. Variability in gene expression by week

The periodicity of peripheral blood gene expression from week to week over the course of one month was evaluated. For all comparisons, the concordance coefficient was high (0.97–0.99) indicating minimal variation in the expression of the majority of genes from week to week. Less than 1% of the expressed genes (5/874) varied more than four-fold over the 4 week period within any one individual (Fig. 3 A–C). Expression of four genes (all from the same male subject) var-

ied significantly between weeks (t-test, $p < 0.02$ with correction for multiple comparisons). As was observed in the time of day comparisons, none of these genes showed a four-fold difference in expression.

4. Conclusions

The peripheral blood is a complex sample that, by nature of circulation through the body, likely reflects many systemic processes. It is also readily collected both in the clinic or field situations making it a good sample for epidemiology studies. These characteristics make peripheral blood ideally suited for expression profiling of diseases that have no known or accessible lesion. In some studies, however, it is not always possible to control when a peripheral blood sample is collected. While much is known about the expression variability of genes related to circadian and hormonal rhythms, little is known about the normal variability of a broad spectrum of genes from a variety of biological pathways. In order to differentiate disease-specific differential gene expression from normal peripheral blood expression variation, it was necessary to understand how much variation of the 1,176 genes we were interested in testing was due to the time of day the sample was obtained.

We minimized possible effects on gene expression due to sample handling and processing by isolating PBMCs within two hours following venipuncture. In addition, we requested the subjects to not eat three hours prior to venipuncture. Under these conditions no

consistent time variation was noted in the expression of any genes. Less than 1% of the genes showed any variation by either method, and none of the specific changes were noted in more than one individual. These results suggest that observed variation was within the experimental limits and reproducibility of this expression profiling technique [10]. Since the arrays we used contained a broad spectrum of genes representing several functional families, our observation should apply for others using similar array formats.

Peripheral blood is subject to changes in cell population and protein activity in response to many factors including stress, hormone levels, time of day, and sleep schedules [1,5,6,11]. It is important to determine whether these cell population and activity changes are observed at the gene expression level as well. While the subset of known human genes included in this study is widely distributed among various gene families (data not shown), only 1,176 genes were assessed and included a limited number of cytokines and circadian rhythm genes known to exhibit time dependent expression [2,6]. The results of expression profiles containing more genes such as these would likely exhibit expected time dependent differential gene expression.

There are limitations in our exploratory study. We studied only three individuals over a limited time course (three times a day once a week for one month) and used only one array that includes 1,176 genes. These results provide data required for study design of a more robust analysis of this question that would include more age and sex matched subjects sampled over a longer time frame with different arrays. In summary, our study finds that the peripheral blood gene expression of a broad spectrum of known function genes printed on the 1.2 II Atlas arrays exhibit little normal variability related to the time of sample collection. Thus, existing and carefully processed PBMC samples collected without systematic control for time of collection could be used in array-based gene expression studies to identify differentially expressed genes. As with all array experiments, validation of differential gene expression identified with this approach is essential [8].

Acknowledgments

We would like to thank William C. Reeves, M.D., for guidance and helpful discussions throughout this project and Ms. Daisy Lee for her assistance with sample collection and cell preservation techniques. We would also like to thank Ms. Irina Dimulescu for her

help with sample collection and nucleic acid extraction techniques. This research was supported in part by an appointment to the Research Participation program at the Centers for Disease Control and Prevention, National Center for Infectious Diseases, Division of Viral and Rickettsial Diseases administered by the Oak Ridge Institute for Science and Education through and interagency agreement between the U.S. Department of Energy and CDC. Human experimentation guidelines of the U.S. Department of Health and Human Services were followed in the conduct of this study. All study participants were volunteers who gave informed consent.

References

- [1] D.F. Dinges, S.D. Douglas, L. Zaugg, D.E. Campbell, J.M. McMann, W.G. Whitehouse, E.C. Orne, S.C. Kapoor, E. Icaza and M.T. Orne, Leukocytosis and natural killer cell function parallel neurobehavioral fatigue induced by 64 hours of sleep deprivation, *J. Clin. Invest.* **93** (1994), 930–1939.
- [2] P.D. Etter and M. Ramaswami, *The ups and downs of daily life: Profiling circadian gene expression in Drosophila*. *Bioessays* **24** (2002), 494–498.
- [3] P.R. Hedge, Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J.E. Hughes, E. Snesrud, N. Lee and J. Quackenbush, A Concise Guide to cDNA Microarray Analysis, *BioTechniques* **29** (2000), 548–562.
- [4] L. Lin, A concordance coefficient to evaluate reproducibility, *Biometrics* **45** (1989), 255–268.
- [5] A.L. Northern, S.M. Rutter, and C.M. Peterson, Cyclic changes in the concentrations of peripheral blood immune cells during the normal menstrual cycle, *Proc. Soc. Exp. Biol. Med.* **207** (1994), 81–88.
- [6] B. Plytycz and R. Seljelid, Rhythms of immunity, *Arch. Immunol. Ther. Exp.* **45** (1997), 157–162.
- [7] M. Rajeevan, I. Dimulescu, E.R. Unger and S.D. Vernon, Chemiluminescent analysis of gene expression on high density filter arrays, *J. Histochem. Cytochem.* **47** (1999), 337–342.
- [8] M.S. Rajeevan, S.D. Vernon, N. Taysavang and E.R. Unger, Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR, *J. Mol. Diagn.* **3** (2001), 26–31.
- [9] C. Stratowa, G. Loffler, P. Lichter, S. Stilgenbauer, P. Haberl, N. Schweifer, H. Dohner and K.K. Wilgenbus, CDNA microarray gene expression analysis of B-cell chronic lymphocytic leukemia proposes potential new prognostic markers involved in lymphocyte trafficking, *Int. J. Cancer* **91** (2001), 474–480.
- [10] S.D. Vernon, E.R. Unger, M. Rajeevan, I.M. Dimulescu, R. Nisenbaum and C.E. Campbell, Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays, *J. Mol. Diagn.* **2** (2000), 124–127.
- [11] T. Wang, D.L. Delahanty, A.L. Dougall and A. Baum, Responses of natural killer cell activity to acute laboratory stressors in healthy men at different times of day, *Health Psychol.* **17** (1998), 428–435.
- [12] P.H. Westfall and S.S. Young, *Resampling-based multiple testing*, John Wiley and Sons, Inc., New York, 1993.

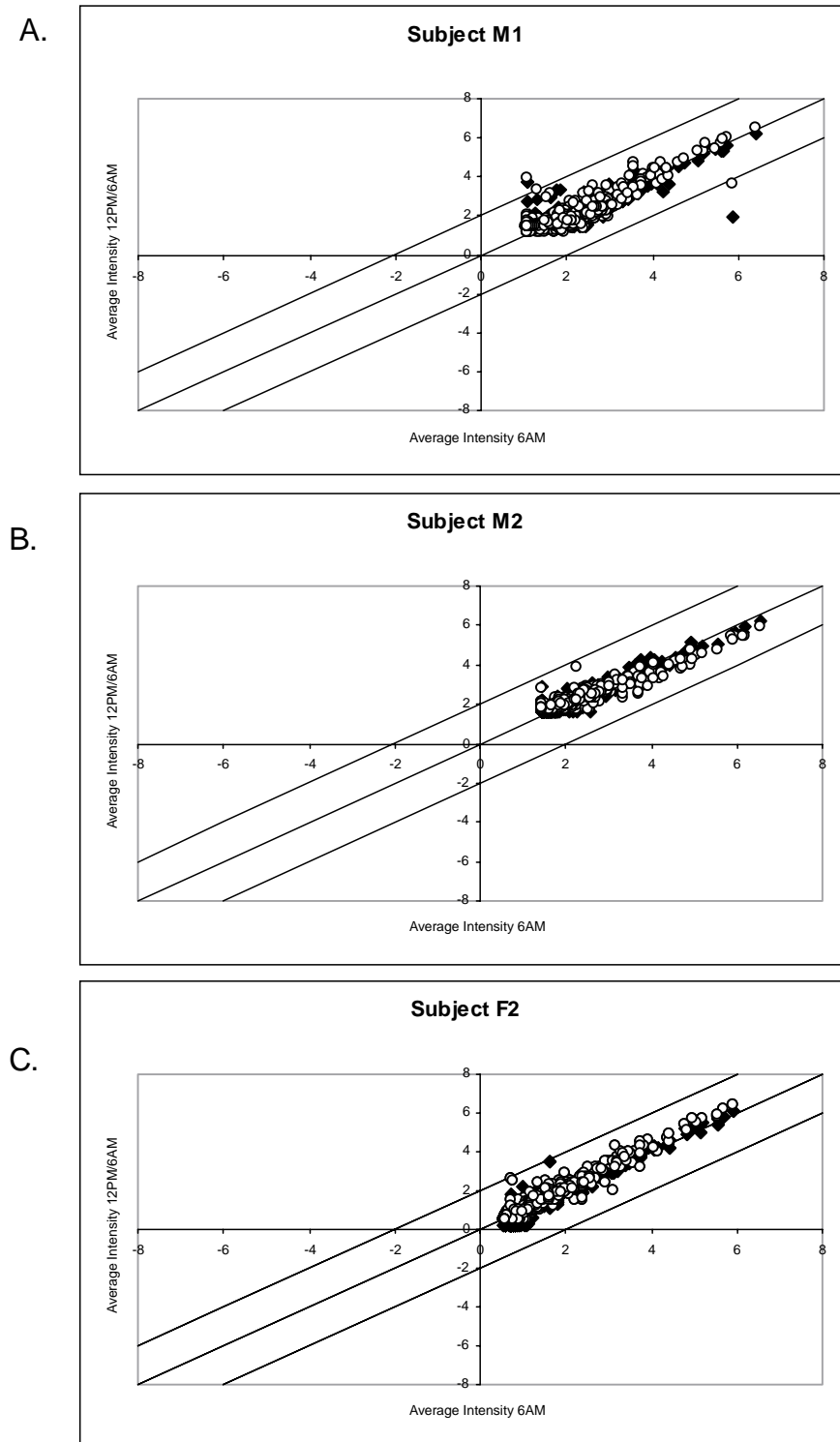


Fig. 2. Variability in intensity of gene expression by time of day in each of three subjects. These three graphs (A,B, and C) show the variability in gene intensities in each of three subjects comparing 6 a.m. to 12 p.m. and 6 a.m. to 6 p.m. Weeks 1 through 4 have been averaged for each time point. The line through the center of the graph is a 1:1 identity line, and the two outer diagonal lines mark four-fold differences in gene intensity. Less than 0.5 % of genes vary by more than four-fold for each individual.

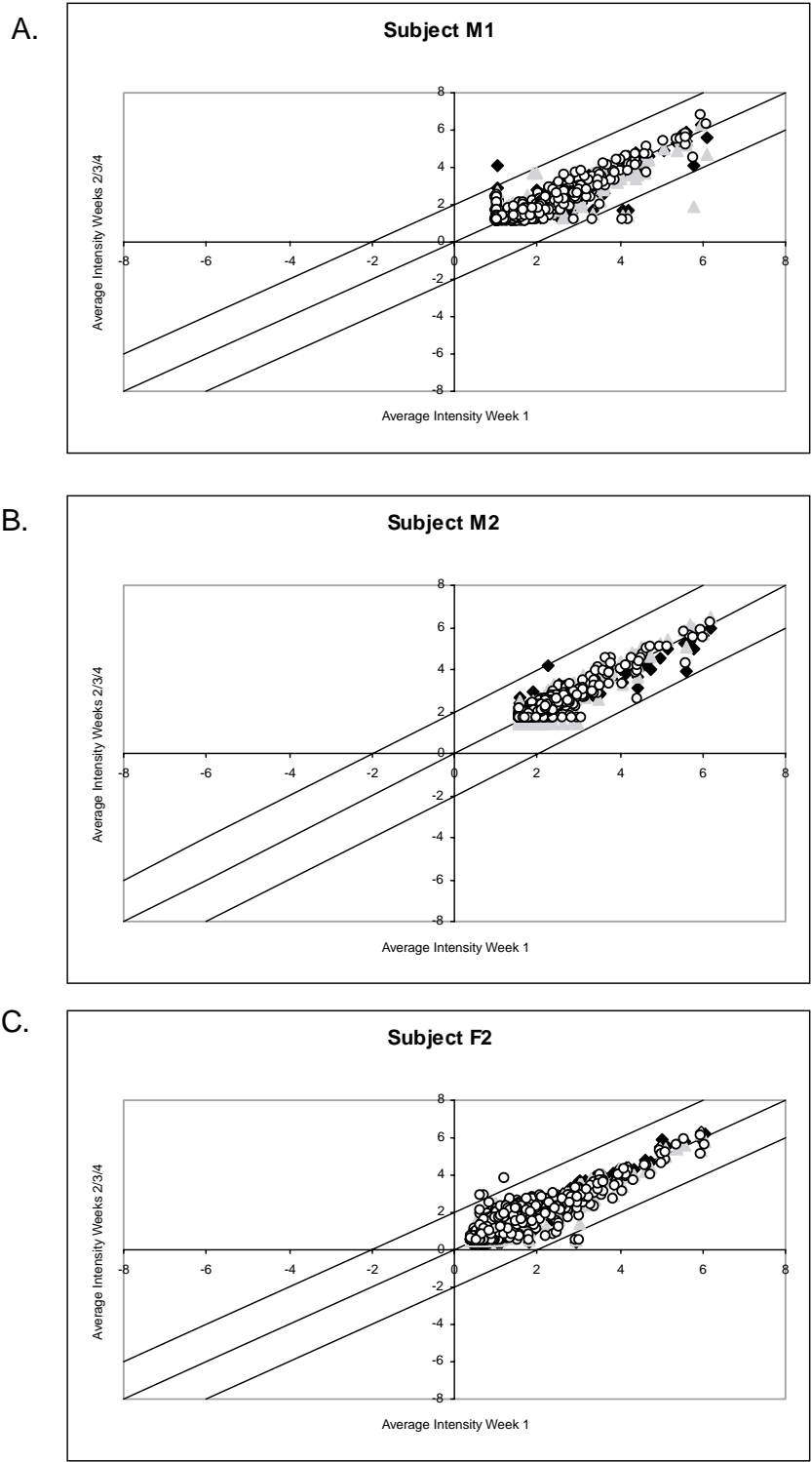


Fig. 3. Variability in intensity of gene expression by week in each of three subjects. These three graphs (A,B, and C) show the variability in gene intensities in each of three subjects comparing Week 1 to Weeks 2, 3, and 4. Time points 6 a.m., 12 p.m., and 6 p.m. have been averaged for each week. The line through the center of the graph is a 1:1 identity line, and the two outer diagonal lines mark four-fold differences in gene intensity. Less than 1 % of genes vary by more than four-fold.