Nanoliter high-throughput RT-qPCR: a statistical analysis and assessment

James M. Dixon[1], Mariusz Lubomirski[1], Dhammika Amaratunga[1], Tom B. Morrison[2], Colin J.H. Brenan[2], and Sergey E. Ilyin[3]
[1]Johnson & Johnson Pharmaceutical Research & Development, L.L.C., Spring House, PA, USA, [2]BioTrove, Inc., Woburn, MA, USA, and [3]Serigene L.L.C., Doylestown, PA, USA

BioTechniques 46:ii-viii (*BioTechniques for Preclinical Development* May 2009) doi 10.2144/000112746
Keywords: high-throughput real-time PCR; nanoliter PCR assessment; kinase gene expression; spectral maps; nanoliter fluidic systems

Biomarkers discovered from gene expression profiles using hybridization microarrays have made great inroads in the diagnosis and development of safer and efficacious drugs. The candidate gene set is biologically validated by quantitative measurement with reverse transcriptase quantitative PCR (RT-qPCR) and is an effective strategy when implemented with microplates if the number of candidate genes and samples is small. With the trend toward informative candidate gene panels increasing from tens to hundreds of genes and sample cohorts exceeding several hundred, an alternative fluidic approach is needed that preserves the intrinsic analytical precision, large dynamic range, and high sensitivity of RT-qPCR, yet is scalable to high throughputs. We have evaluated the performance of a nanoliter fluidic system that enables up to 3072 nanoliter RT-qPCR assays simultaneously in a high-density array format. We measured the transcription from two different adult human tissues to assess measurement reproducibility across replicates, measurement accuracy, precision, specificity, and sensitivity; determined the false positive rate (FPR) and false negative rate (FNR) of the expressed transcript copies; and determined differences in kinase gene expression reflecting tissue and dosage differences. Using our methodology, we confirm the potential of this technology in advancing pharmaceutical research and development.

Introduction

Quantified measurement of gene transcription is critical to elucidate the mechanisms of cell survival, growth, and differentiation in homeostasis and disease. Oligonucleotide (1,2) and cDNA (3) hybridization microarrays have emerged as the leading analytical tools for de novo discovery of gene expression patterns, due to their ability to record the transcribed messages from many thousands to tens of thousands of genes in a sample simultaneously (4). Typically, only the activity of a subset of genes is of interest in answering a specific biological question; this hypothesis is validated by screening the selected gene set against a larger sample population representing greater biological diversity, or set of test conditions. This secondary screen is performed to overcome inherent microarray deficiencies (5–9) and is implemented as a solution phase, real-time reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) assay in 96- or 384-well microplates. Increases in wait time, labor, and reagent consumption, as well as a more-complex workflow, are fundamental limitations to a wider adoption of RT-qPCR as

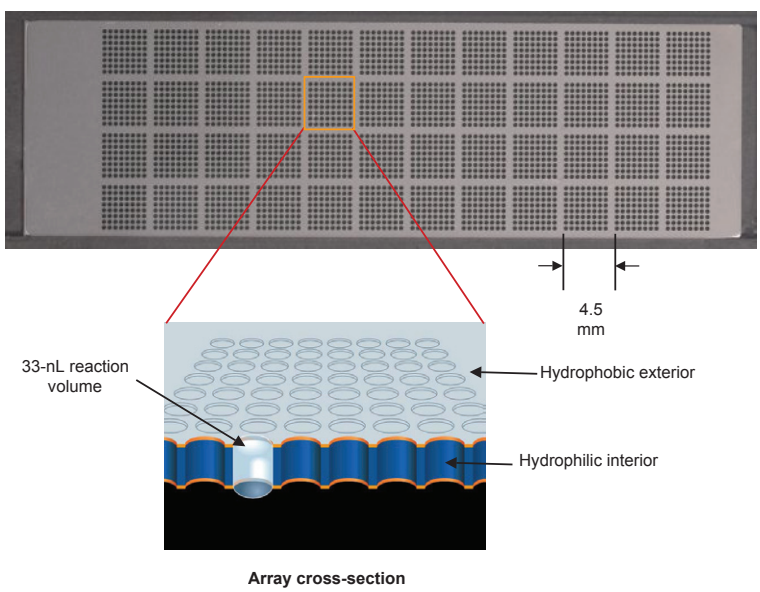

Figure 1. Photograph of the nanofluidic through-hole array with a cartoon depicting the device cross-section. Polymeric coatings are covalently linked to the etched stainless steel platen to make the inside surface of each through-hole hydrophilic and the exterior surface of the platen hydrophobic. The differential surface coating generates a capillary force that retains fluid in each through-hole in isolation from its neighbors.

a primary or secondary screening tool for large sample sizes.

There is, therefore, a need for systems capable of implementing high-throughput RT-qPCR with a simplified workflow and reduced reagent consumption to make scale-up possible for screening the activity of tens to hundreds of genes simultaneously in hundreds to thousands of biological samples. Integral to the development and application of these systems are the methodologies for performance assessment based on standard system metrics of measurement reproducibility, accuracy, precision, sensitivity, and specificity (10,11). Since multiple values are measured simultaneously, it is important to characterize the capabilities of the system for quantitative measurement of gene expression based on such factors as signal magnitude and variability.

Table 1. Differentially Expressed Genes

GenBank Accession No.	Gene Name
NM_003160	Aurora kinase C (AURKC), transcript variant 3
NM_001824	Creatine kinase, muscle (CKM)
NM_018208	Ethanolamine kinase 2 (ETNK2)
NM_003614	Galanin receptor 3 (GALR3)
NM_014365	Heat shock 22 kDa protein 8 (HSPB8)
NM_000221	Ketohexokinase (fructokinase) (KHK), transcript variant A
NM_006488	Ketohexokinase (fructokinase) (KHK), transcript variant B
NM_018650	MAP/microtubule affinity-regulating kinase 1 (MARK1)
NM_000914	Opioid receptor, Mu 1 (OPRM1), transcript variant MOR-1
NM_138316	Pantothenate kinase 1 (PANK1), transcript variant gamma
NM_148978	Pantothenate kinase 1 (PANK1), transcript variant beta
NM_002591	Phosphoenolpyruvate carboxykinase 1 (soluble) (PCK1)
NM_002625	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1 (PFKFB1)
NM_003558	Phosphatidylinositol-4-phosphate 5-kinase, type I, beta (PIP5K1B), transcript variant 2
NM_000298	Pyruvate kinase, liver and RBC (PKLR), nuclear gene encoding mitochondrial protein, transcript variant 1
NM_002692	Polymerase (DNA directed), epsilon 2 (p59 subunit) (POLE2)
NM_006252	Protein kinase, AMP-activated, Alpha 2 catalytic subunit (PRKAA2)
NM_002732	Protein kinase, cAMP-dependent, catalytic, gamma (PRKACG)
NM_006648	WNK lysine deficient protein kinase 2 (WNK2)
NM_020630	Ret proto-oncogene (multiple endocrine neoplasia and medullary thyroid carcinoma 1) (RET), transcript variant 4
NM_170693	Serum/glucocorticoid regulated kinase 2 (SGK2), transcript variant 1
NM_014978	Sortilin-related VPS10 domain containing receptor 3 (SORCS3)
NM_003280	Troponin C type 1 (slow) (TNNC1)
NM_005108	Xylulokinase homolog (H. influenzae) (XYLB)
NM_001715*	B lymphoid tyrosine kinase (BLK)
NM_006549*	Calcium/calmodulin-dependent protein kinase 2, β (CAMKK2), transcript variant 2
NM_000730*	Cholecystokinin A receptor (CCKAR)
NM_001825*	Creatine kinase,, mitochondrial 2 (CKMT2)
NM_152247*	Carnitine palmitoyltransferase 1B (CPT1B)
NM_004119*	FMS-related tyrosine kinase 3 (FLT3)
NM_053029*	Myosin, light polypeptide kinase (MYLK), transcript variant 4
NM_004563*	Phosphoenolpyruvate carboxykinase 2 (mitochondrial) (PCK2), Nuclear gene encoding mitochondrial protein
NM_000289*	Phosphofructokinase, muscle (PFKM)
NM_002627*	Phosphofructokinase, platelet (PFKP)
NM_006201*	PCTAIRE Protein kinase 1 (PCTK1), transcript variant 1

List of 35 most distal genes from the spectral map separating two tissues. The 11 genes marked with asterisks (shown as blue circles on the map in Figure 9) are known to be differentially expressed in the two tissues.

Miniaturized systems for high-throughput quantitative PCR

Quantitative or real-time PCR (12) is a derivative of the conventional PCR process (13) wherein the fluorescent signal is recorded at the primer annealing temperature in each temperature cycle. This signal is proportional to the number of DNA or cDNA template copies in the sample and is parameterized by defining the cycle threshold (Ct) as the temperature cycle at which the fluorescent signal is at least three standard deviations above the mean background fluorescence. Calibration by way of a standard curve, which relates Ct to number of transcript copies for a specific gene, makes determining its copy number in a sample possible. RNA quantification requires reverse transcription of RNA into cDNA prior to application of the real-time PCR method.

Strategies for increasing the analytical throughput of RT-qPCR involve either automating the manual workflow developed for microplates, multiplexing several RT-qPCR measurements in a single reaction, or decreasing the PCR assay volume to the microliter- or nanoliter-scale and increasing throughput by either fast serial or highly parallel processing of each miniaturized reaction. Automation of the RT-qPCR process increases the number of genes and samples that can be analyzed and improves the uniformity of data quality compared with a manual workflow, but these gains are offset by increased sample and reagent consumption, capital equipment costs, and the need for specialized labor to run and maintain robotic equipment (14,15). Multiplexed RT-qPCR analyses allow the number of transcript copies for multiple genes to be measured simultaneously in a single sample (16). However, it is a significant challenge to design primer pairs with a high and reproducible specificity to the target sequence and a measurable and reproducible amplification efficiency for each transcript targeted (17).

Miniaturization of PCR volumes without sacrificing data quality allows the number of analyses to increase without consuming more reagents and enables implementation of serial or parallel processing strategies to achieve high throughput. Most reports of microliter, nanoliter or picoliter PCR are for sequence-specific detection and not measurement of the number of expressed gene copies. Continuous-flow PCR devices utilize etched microchannels with fixed temperature zones to reduce reaction volumes to sub-microliter levels and analyze PCR products by hybridization followed by electrochemical

detection (18), fluorescence detection (19), or electrophoretic separation with fluorescent detection (20). However, a key drawback in continuous-flow systems is the limited sample throughput and potential for cross-contamination that can result from processing and analyzing samples through a common microchannel. Devices for parallel PCR processing involve thermal cycling PCR reagents and sample template in a high-density array (>1 well/mm^2) of micro-, nano- or even picoliter wells. Quantification of transcript copy number for a specific gene in these devices uses endpoint fluorescence of a probe compared with a standard curve (21,22) or digital PCR (23) based on template dilution and counting the number of positive assays for a specific target template (24). The distribution of template molecules at low concentration follows Poisson statistics; therefore, high-density arrays of wells are well-suited for high dynamic range measurement down to single-template copies from partitioning of the dilute sample into many thousands of individual picoliter containers (25–27). Despite its utility, the dynamic range of digital PCR is limited by the number of wells in the array and the accuracy of sample dilution.

A preferred practice is the implementation of the real-time PCR method for microplates but in a miniaturized, high-density array format. A statistics-based methodology similar to that used to evaluate hybridization microarrays is necessary for evaluating the performance of such high-throughput real-time PCR systems. Evaluation metrics include the assessment of real-time PCR data quality based on array-to-array reproducibility, precision and accuracy in measuring differences in gene expression between samples, and a gene-by-gene estimate of false positive and false negative errors as a measure of copy number sensitivity and specificity of amplification. Our evaluation by statistical analysis of gene expression data produced by the system described here confirms and supports the potential for this technology in drug discovery and disease research applications and the overall appropriateness of the proposed analytical methodology.

Materials and methods

Technical specifications

The nanoliter fluidic system (OpenArray NT Cycler System and Human Kinome Panel Kit; BioTrove, Inc., Woburn, MA USA) evaluated in this study has been described previously (28–31) and is based on a high-density, rectilinear array of 3072 etched through-holes in a thin stainless steel platen (25 mm × 75 mm × 0.3 mm) with polymer coatings engineered to make the inside surface of each hole hydrophilic and the exterior surface of the plate hydrophobic. The through-holes are grouped into 48 subarrays of 64 holes each, and spaced on a 4.5-mm pitch equal to that of wells in a 384-well microplate. Each 33-nL hole in the nanotiter plate is a separate and fluidically isolated container into which primer pairs are transferred in such a way that each through-hole in the plate can support a different PCR assay

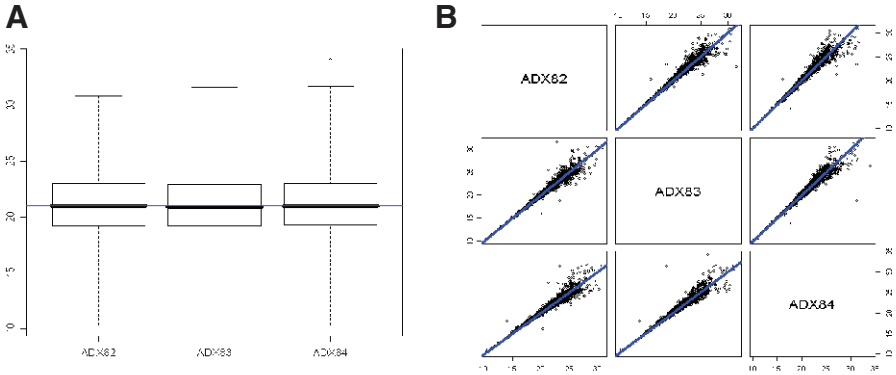

Figure 2. Inter-plate reproducibility. (A) Box plots and pairs plots of the raw data from 3 replicate chips ADX82, ADX83, and ADX84. x-axis, plate number; y-axis, Ct. (B) Spearman correlation coefficient across chips is >0.985, and concordance correlation coefficient is >0.965. Each demonstrates high consistency of data distribution across nanoplates x-axis, y-axis, Ct.

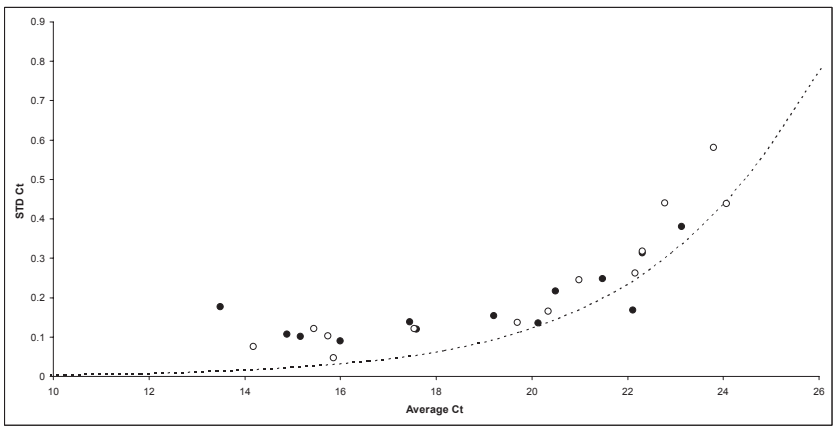

Figure 3. Comparison of the cycle threshold (Ct) standard deviation (STD) against the average Ct based on 16 replicates of 13 housekeeping genes for the heart (open circle) and liver (closed circle) samples. Ct variance at Ct>18 (low copy number) follows a model based on Poisson statistics (dotted line). At an average Ct<18 (high copy number) the Ct standard deviation reaches a constant value of approximately 0.1, indicative of a noise process independent of sample copy number.

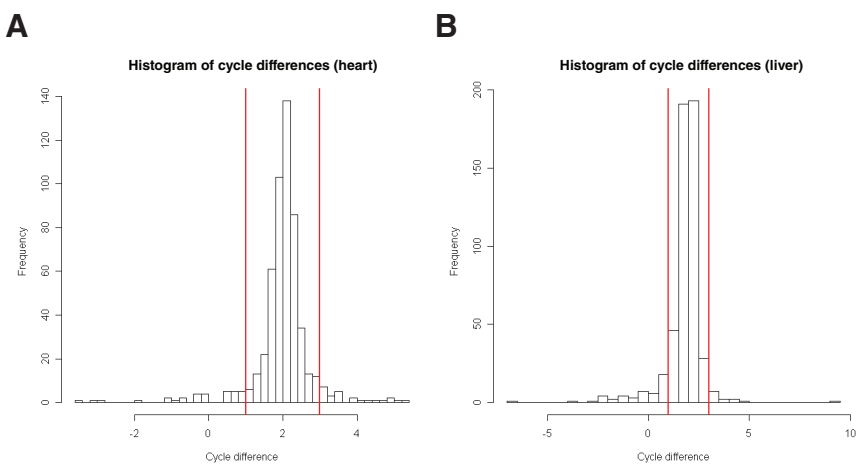

Figure 4. Histogram of median cycle differences between 8 ng/µL and 32 ng/µL total RNA dosages of human heart and liver tissues. (A) is heart tissue and (B) is liver tissue, with 3 replicates each. The center difference is 2 cycles with vertical lines showing ±1 cycle interval.

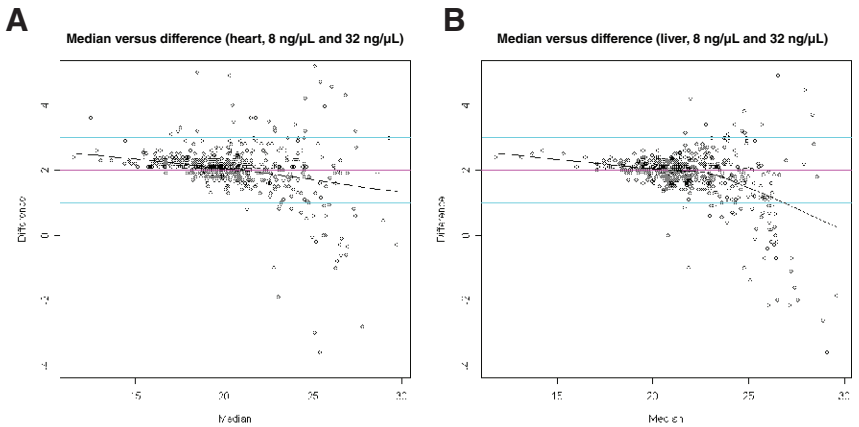

Figure 5. Cycle differences plotted against cycle medians. Cycle medians plotted for both (A) heart and (B) liver tissue confirm the expected 2-cycle difference but also reveal bias. The black line running through the points is a lowess smooth (39) of the data. The blue lines mark ±1 cycle deviation. x-axis, median Ct; y-axis, cycled difference.

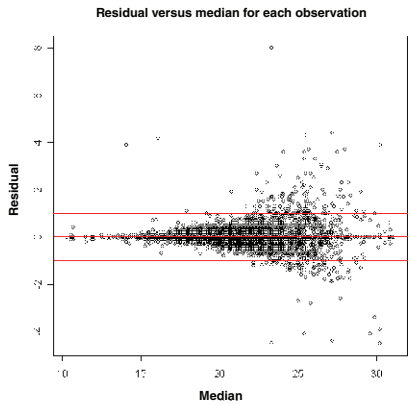

Figure 6. Residual versus median of cycles for all observations from all chips. Observation is a number of cycles for a single gene on any one chip. x-axis, median Ct; y-axis, cycled difference. Median is calculated as a median of observations for a single gene taken across three replicates (chips). Residuals are calculated as a difference between single gene expression cycles across three replicates (chips) and its median. Red lines indicate ±1 cycle deviation.

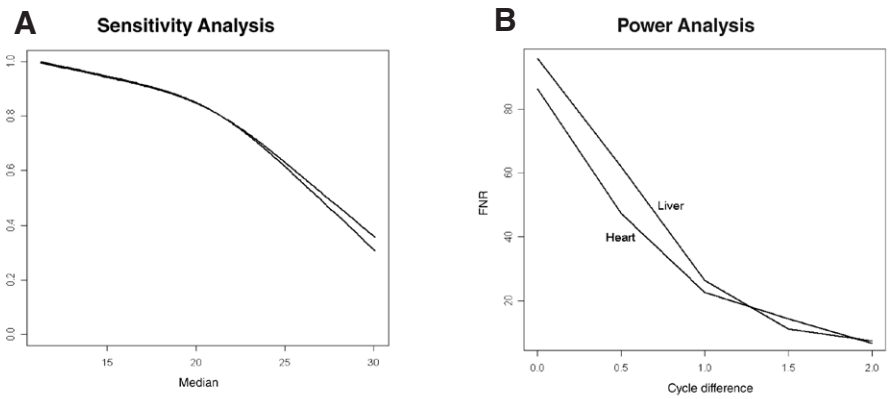

Figure 7. Sensitivity and false-negative rates for the human heart and liver gene sxpression data. (A) Sensitivity plots for the two types of tissues. Heart tissue is represented by the upper curve, and liver tissue is represented by the lower curve. x-axis, Ct; y-axis, 1-FNR. (B) False negative rate (FNR) increases as a function of decreasing cycle difference within each tissue under investigation. x-axis, Ct; y-axis, FNR.

(Figure 1). Once a platen is fully populated with primer pairs, the solvent is evaporated in a controlled manner, leaving the primers immobilized in a PEG matrix on the inside surface of each through-hole. Previously prepared cDNA samples at a final concentration of 32 ng/µL are mixed with off-the-shelf RT-qPCR reagents [SYBR Green I kit (Roche Applied Science Indianapolis, IN, USA), Exonuclease I (GE Healthcare, Piscataway, NJ, USA), Pluronic F-68 (Invitrogen, Carlsbad, CA, USA), BSA (Sigma-Aldrich, St. Louis, MO, USA), SYBR Green I (Sigma-Aldrich), 99% glycerol (Sigma-Aldrich), Formamide (Sigma-Aldrich)] for SYBR Green PCR and dispensed into each sub-array with an automated pipette tip dispensing device (OpenArray Autoloader, BioTrove, Inc.). The sample assay layout is flexible, allowing 48 samples to be interrogated by 64 PCR assays per sample all the way to one sample queried by 3072 PCR assays. Self-metered filling of each through-hole is driven by capillary forces and the difference in surface energies between the hydrophobic and hydrophilic surfaces. Volume precision and accuracy is high (CV < 2%) and is determined by the uniformity of hydrophilic/hydrophobic polymer coating deposition and the precision and accuracy of the micromachined through-hole dimensions.

The prepared platen is inserted into a glass-walled, slotted cassette containing an immiscible perfluorinated liquid (Fluorinert; 3M, St. Paul, MN, USA) to prevent evaporation during thermal cycling and hermetically sealed for thermal cycling by UV-cured epoxy. The cassettes are placed in a computer-controlled imaging thermal cycler instrument capable of implementing 9216 real-time PCR amplifications and dissociation curves in <4 h. Post-acquisition data processing generates fluorescence amplification and melt curves for each through-hole in the array, from which cycle threshold (Ct) and melt temperature (T_m) are computed for each reaction. All data are stored in a comma-separated value (.csv) flat file format for simple export to a database or third-party software for further analysis.

Experiment

Primer pairs targeting 508 human kinase genes and 13 housekeeping genes as endogenous controls were designed and validated according to the method described in Morrison et al. (31). The primer sets were loaded into a nanotiter plate such that each sample is interrogated by 508 kinase gene assays, 13 housekeeping gene assays in quadruplicate, and 208 negative (no assay) controls, allowing 4 samples to be processed per nanotiter plate. The negative control was designed to detect interhole primer carryover, of which none was observed. To assess the system performance, normal human adult liver and heart total RNA samples (Cat nos. 636531 and 636532, respectively; Clontech, Mountain View, CA, USA) were converted by reverse transcriptase into cDNA and prepared at two different final concentrations (8 ng/µL and 32 ng/µL). These were dispensed into the nanotiter plate to measure target gene expression in triplicate or three identical chip layouts to inspect the precision of each measurement and ensure data reliability. Liver and heart total RNA samples were chosen as the validation samples because their expression profiles are well known and both have a number of genes that differentially express. The use of two different, but known, concentrations of mRNA for each tissue allowed us to verify that the technology is capable of detecting such a difference. This then serves as a scheme to assess the measurement bias of the system, which cannot be measured directly. The data recorded by the instrument was analyzed

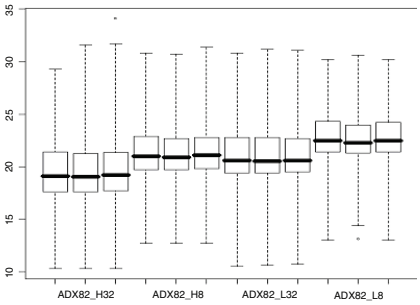

Figure 8. Box plots of raw data from the chip ADX82 of individual tissues. Plot shows data from heart and lung tissues at 8 ng/µL (ADX82_H8 and ADX82_L8, respectively) and 32 ng/µL ADX82_H32 and ADX82_L32, respectively) dosages. y-axis, Ct. Data from chips ADX83 and ADX84 showed a similar pattern.

using the R software platform (Comprehensive R Archive Network, cran.r-project.org), with the use of the DNAMR library modules (www.rci.rutgers.edu/~cabrera/DNAMR).

Results

Consistency across nanotiter plates

A common method for evaluating consistency across microarrays is examination using box and pairs plots of the raw Ct data (Figure 2). For the nanotiter plate RT-qPCR data, the plots showed highly consistent results among the nanotiter plate replicates. Agreement among nanoplates is particularly noticeable for low Ct values (high transcript abundance) whereas high Ct values (low transcript abundance) display more variability, but the average agreement is maintained. The pairwise Spearman correlation coefficients across chips are all >0.985 and the concordance correlation coefficients, which measures linear dependence at tangent equal to one, are all >0.965.

To determine the line between system noise (hole volume, Ct value calculation, thermal uniformity, etc.) and starting copy number Poisson statistics, we examined the 16 technical replicates for each of the housekeeping genes spread over 3 arrays. Figure 3 compares the Ct value standard deviations (STDs) to the average Ct values across the 3 replicates. The data show that Monte Carlo errors (32) affect precision at lower starting copies (Ct<18). However, Ct standard deviation reaches a constant value of ~0.1 for Ct<18, indicative of a noise process independent of sample copy number.

These findings indicate that it is not necessary to carry out data normalization for relative expression analysis, which is

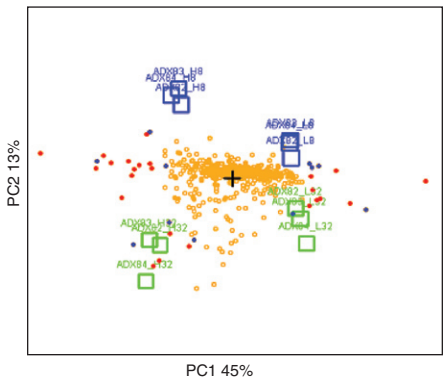

Figure 9. Spectral map showing separation of heart and liver tissues along the first principal component. Heart tissue is represented on the left of the map as blue (8 ng/µL) and green (32 ng/µL) squares. Liver tissue is represented on the right of the map as blue (8 ng/µL) and green (32 ng/µL) squares. The horizontal separation is due to the tissue differences while vertical separation is due to the dosage differences. Circles represent individual observations or genes; red and blue represent the most distal genes in the horizontal direction, and thus the most influential in the separation of heart and liver tissues. Genes marked in blue are well-described in literature to be differentially expressed between the two tissues types. Genes marked in red are identified by the map as differentially expressed but require further validation. PC, principal component.

something that would normally be required for a different type of approach (i.e., high-throughput microarray data). This is hardly surprising since not only are the fluorescent signals proportional to the number of hybridized fragments in a microarray; they are also proportional to the length of these fragments and the number of fluorescent labels each fragment carries (33). This is not the case with RT-qPCR, thus eliminating the need for data normalization.

Accuracy of differential expression

Most RT-qPCR experiments are comparative in nature, with an objective to estimate the level of differential expression of a series of genes across two or more conditions. As such, rather than assess the accuracy of an individual measurement, we assessed the accuracy of an estimate in fold change. A four-fold difference in concentration should provide a 2-cycle difference between the corresponding tissues. To verify that the system can detect this, we calculated the cycle differences between the two concentrations for liver and heart tissues, respectively. Figure 4 shows the distribution of the resulting values with the histogram median for both heart and liver equal to the 2-cycle difference with 99.9% of the variance within ±1 cycle of the median difference. The ±1 cycle variability is due to the intrinsic distribution of the copy number for each assay template in the sample. Within this population, the statistical variation of copy number will be replicated in the statistical variation of cycle number as observed in Figure 4. The observed variation is higher than in larger-volume RT-qPCR, but this is expected as with smaller volumes comes a larger Poisson effect and Monte Carlo errors (32).

To determine whether there was any bias associated with Ct level, we also examined the plot of differences in cycles between the two dosages versus the median values of the measured observations, Figure 5. The difference of 2 ± 1 cycles is again clearly demonstrated, but a small but consistent bias with negative slope is present in the data, resulting in, on average, a difference of >2 cycles at low median cycles and a difference of <2 cycles at high median cycles. Nevertheless, for genes with low median Ct values, most differences are within a cycle of the expected 2-cycle difference. On the other hand, the increased variance from Poisson noise for quite a few of the genes with high median Ct values results in a cycle difference substantially different from two. With Ct≤22 cycles, 95% of all observations fall within ±1 cycle. This is not surprising, since a standard curve measurement based on template dilution showed that Ct>25 corresponds to a transcript copy number of <1 (31).

Precision of differential expression

The system precision across the entire range of data cycles was examined to see whether the measurement variability is a function of the amount of mRNA starting copy number. A scatterplot of residuals versus their corresponding medians for each gene is shown in Figure 6. The average CV of the measurements was an acceptable 1.52% (0.88%; Ct≤22 cycles) for heart tissue and 1.65% (0.80%; Ct≤22 cycles) for liver tissue. Similar to Figure 2, an increase in Ct variability is observed with increasing cycle number, indicative of increasing Poisson noise with decreasing target sequence copy number (31). The agreement across replicates for low median Ct values (high copy number) is remarkably strong with very few residuals exceeding a cycle.

Analysis of sensitivity and specificity

Sensitivity and specificity rates are excellent markers for determining the potential utility of an instrument in a particular application. In our experiment, we examined both rates to see how they change over the range of the

median data cycles and cycle differences. Our sensitivity analysis used the conditional *t*-test (34,35) via a gene-by-gene approach to detect differences and subsequently calculate positive and negative discovery rates. We report these rates with and without adjustment for bias, which was shown to exist (Figure 5). The false positive rate (FPR) for heart tissue was 8.8% (11.9% with bias) and 1.5% (3.2% with bias) for liver tissue. The false negative rate (FNR) was 6.2% (6.6% with bias) for heart and 7.9% (10.6% with bias) for liver. A sensitivity plot of the two tissues, as a function of cycle medians with bias of Figure 5 removed, is shown in Figure 7A. The rates are very similar in the region where precision is high (Ct<22) with a break point indicated by a increased slope at Ct>22. Again, this corresponds to very low copy number and can be expected. The difference in sensitivity with increasing Ct value to above 22 cycles for the two tissue types is consistent with the increased variance between the two tissues.

The ability of the system to correctly and consistently detect a 2-fold difference in cycle number adds an immense amount of statistical power. To show statistical power, we display false negative rate as a function of cycle difference (Figure 7B). The curves are very similar between the two tissues with the most power at a cycle difference of ~2, and decreasing steadily toward a cycle difference of ~1, when a rapid fall in power is observed.

Separation of tissues and dosages

We have already discussed in detail the separation of dosages by examining cycle differences. Since our experiment consists of two different types of tissues, we will now explore whether the purported differences among tissues are indeed observable in the data. A box plot of all replicates for both tissue types at each dose is shown in Figure 8. The first 6 samples in Figure 8 refer to heart tissue at 32 ng/µL and 8 ng/µL respectively; the remaining 6 refer to liver tissue at 32 ng/µL and 8 ng/µL respectively. The 2-cycle differences between the concentrations are visible as illustrated previously, but additionally, there is a shift in cycles between tissues at the same dosage. Relatively little work has been done in genomics to determine the differences in total amounts of mRNA in different normal human tissues. However, in other species such as mouse (36), there is a higher percentage of mRNA from the liver than heart with similar amounts of total RNA. We feel it is safe to assume the observed one-cycle difference is due to tissue differences and not a system bias. Mice are generally products of controlled breeding and environments, and as such, very little sample variation is expected. Humans vary greatly in their genetics and environment, so much larger differences are to be expected.

A more detailed examination of the differences between tissues can be performed using spectral maps (37). The use of spectral maps, which is an excellent method for visualizing multivariate data, is a very important step as it serves as a proof of concept showing that a dominant separating signal is indeed present in the data. Figure 9 shows the spectral map of liver and heart tissues at the two dosages. The first principal component separates the two tissue types while the second principal component separates the dosages. There is enough evidence to suggest that there exists distinguishable features in the data that separate the two sample types and dosages. The most distal genes shown in red and blue (Figure 9) are the genes most highly correlated with the separation of tissues. In fact, the genes shown in blue are very well described in literature and have been compiled into a single database (38) to have differential expression between the two types of tissues. The results of the separation of the two tissues using a spectral map are summarized in Table 1, which contains the list of all of the most distal genes.

Discussion

We have successfully demonstrated the use of statistical methods for analyzing high-throughput RT-qPCR data generated by a novel nanoliter fluidic system. Methods for assessing the quality of the data and for detecting differentially expressing genes have been outlined. A summary of the major findings is as follows: (*i*) the data quality was consistent across the chips and did not require typical microarray style normalization, and (*ii*) the data was generally accurate, with a majority of the data points with cycles <22 lying within ±1 cycle of the target (2 cycles) for differential expression. However, (*iii*) there was a small but clear bias in the data, which persisted throughout the entire range of cycles. The bias resulted in higher readings at low cycle mean values and lower readings at high cycle mean values. Nevertheless, with statistical treatment, the data was easily corrected and provided precise estimates of differential expression. (*iv*) False positive rates and false negative rates, with and without bias, were estimated to provide insight into whether the data analysis was sensitive and specific. The false positive rates showed only a small difference in outcome. Conversely, the false negative rates were found to increase rapidly for cycle differences <1, with negligible impact on data bias. (*v*) The use of spectral maps demonstrated the existence of dominant signals in the data, which separated both tissue samples and dosages. A set of 35 genes was identified to differentiate the tissue samples with a subset of 11 already known as separators. (*vi*) A Ct value difference of 1 between heart and liver samples suggests a difference in total RNA abundance. It is conjectured that this corresponds to a difference in the mRNA abundance, as the total RNA input was constant input.

We have only offered one example, of many possible, for an assessment of utility. Nonetheless, using this set of criteria, we have successfully demonstrated that (*i*) the high-throughput RT-qPCR instrumentation evaluated herein is capable of generating meaningful signal, and (*ii*) statistical methods, like spectral maps, are capable of detecting and differentiating these signals. We believe this statistical analysis applied to RT-qPCR data generated from high-throughput instrumentation could revolutionize searches for targets and biomarkers by enabling the quantitative analysis of the expression of large numbers of genes across large sample cohorts.

Acknowledgments

We would like to thank Elen Ortenberg of BioTrove for processing the OpenArray plates used in this study.

C.B. and T.M. are employed by BioTrove, Inc., which manufactures and sells the OpenArray device and software described in this manuscript.

References

1. **Pease, A.C., D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, and S.P. Fodor.** 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. USA *91*:5022-5026.
2. **Lipshutz, R.J., D. Morris, M. Chee, E. Hubbell, M.J. Kozal, N. Shah, N. Shen, R. Yang, and S.P. Fodor.** 1995. Using oligonucleotide probe arrays to access genetic diversity. BioTechniques *19*:442-447.
3. **Schena, M., D. Shalon, R.W. Davis, and P.O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*:467-470.
4. **Gershon, D.** 2002. Microarray technology: an array of opportunities. Nature *416*:885-891.
5. **Bammler, T., R.P. Beyer, S. Bhattacharya, G.A. Boorman, A. Boyles, B.U. Bradford, R.E. Bumgarner, P.R. Bushel, et al.** 2005.

Standardizing global gene expression analysis between laboratories and across platforms. Nat. Methods *2*:351-356.
6. **Murphy, D.** 2002. Gene expression studies using microarrays: Principles, problems and prospects. Adv. Physiol. Educ. *26*:256-270.
7. **Petersen, D., G.V. Chandramouli, J. Geoghegan, J. Hilburn, J. Paarlberg, C.H. Kim, D. Munroe, L. Gangi, et al.** 2005. Three microarray platforms: an analysis of their concordance in profiling gene expression. BMC Genomics *6*:63.
8. **Canales, R.D., Y. Luo, J.C. Willey, B. Austermiller, C.C. Barbacioru, C. Boysen, K. Hunkapiller, R.V. Jensen, et al.** 2006. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat. Biotechnol. *24*:1115-1122.
9. **Larkin, J.E., B.C. Frank, H. Gavras, R. Sultana, and J. Quackenbush.** 2005. Independence and reproducibility across microarray platforms. Nat. Methods *2*:337-344.
10. **Allchin, D.** 2001. Error types. Perspect. Sci. *9*:38-58.
11. **Ji, H. and R.W. Davis.** 2006. Data quality in genomics and microarrays. Nat. Biotechnol. *24*:1112-1113.
12. **Higuchi, R., C. Fockler, G. Dollinger, and R. Watson.** 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology (N. Y.) *11*:1026-1030.
13. **Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich.** 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. 1986. Cold Spring Harb. Symp. Quant. Biol. *51*:263-273.
14. **Beuselinck, K., M. van Ranst, and J. van Eldere.** 2005. Automated extraction of viral-pathogen RNA and DNA for high-throughput quantitative real-time PCR. J. Clin. Microbiol. *43*:5541-5546.
15. **Gunson, R.N., T.C. Collins, and W.F. Carman.** 2006. Practical experience of high throughput real time PCR in the routine diagnostic virology setting. J. Clin. Virol. *35*:355-367.
16. **Balashov, S.V., R. Gardiner, S. Park, and D.S. Perlin.** 2005. Rapid, high-throughput, multiplex, real-time PCR for identification of mutations in the *cyp51A* gene of *Aspergillus fumigatus* that confer resistance to itraconazole. J. Clin. Microbiol. *43*:214-222.
17. **Bustin, S.A., V. Benes, T. Nolan, and M.W. Pfaffl.** 2005. Quantitative real-time RT-PCR--a perspective. J. Mol. Endocrinol. *34*:597-601.
18. **Liu, R.H., J. Yang, R. Lenigk, J. Bonanno, and P. Grodzinski.** 2004. Self-contained, fully integrated biochip for sample preparation, polymerase chain reaction amplification, and DNA microarray detection. Anal. Chem. *76*:1824-1831.
19. **Lee, D.S., S.H. Park, H. Yang, K.H. Chung, T.H. Yoon, S.J. Kim, K. Kim, and Y.T. Kim.** 2004. Bulk-micromachined submicroliter-volume PCR chip with very rapid thermal response and low power consumption. Lab Chip *4*:401-407.
20. **Lagally, E.T., I. Medintz, and R.A. Mathies.** 2001. Single-molecule DNA amplification and analysis in an integrated microfluidic device. Anal. Chem. *73*:565-570.
21. **Matsubara, Y., K. Kerman, M. Kobayashi, S. Yamamura, Y. Morita, Y. Takamura, and E. Tamiya.** 2004. On-chip nanoliter-volume multiplex TaqMan polymerase chain reaction from a single copy based on counting fluorescence released microchambers. Anal. Chem. *76*:6434-6439.
22. **Matsubara, Y., K. Kerman, M. Kobayashi, S. Yamamura, Y. Morita, and E. Tamiya.** 2005. Microchamber array based DNA quantification and specific sequence detection from a single copy via PCR in nanoliter volumes. Biosens. Bioelectron. *20*:1482-1490.
23. **Vogelstein, B. and K.W. Kinzler.** 1999. Digital PCR. Proc. Natl. Acad. Sci. USA *96*:9236-9241.
24. **Kalinina, O., I. Lebedeva, J. Brown, and J. Silver.** 1997. Nanoliter scale PCR with TaqMan detection. Nucleic Acids Res. *25*:1999-2004.
25. **Fan, H.C. and S.R. Quake.** 2007. Detection of aneuploidy with digital polymerase chain reaction. Anal. Chem. *79*:7576-7579.
26. **Ottesen, E.A., J.W. Hong, S.R. Quake, and J.R. Leadbetter.** 2006. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. Science *314*:1464-1467.
27. **Warren, L., D. Bryder, I.L. Weissman, and S.R. Quake.** 2006. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc. Natl. Acad. Sci. USA *103*:17807-17812.
28. **Brenan, C.J.H., K. Domansky, P. Kurzawski, and L.G. Griffth.** 2000. BioMEMS applied to the development of cell-based bioassay systems, p. 76-87. SPIE Proceedings.
29. **Brenan, C.J.H., T. Morrison, K. Stone, T. Heitner, A. Katz, T.S. Kanigan, and R. Hess.** 2002. A massively parallel microfluidics platform for storage and ultra high throughput screening, p. 560-569. SPIE Proceedings.
30. **Kanigan, T.K., C.J.H. Brenan, S. Lafontaine, L. Sosnowski, P.G. Madden, and I.W. Hunter.** 2000. Living Chips for drug discovery, p. 172-180. SPIE Proceedings.
31. **Morrison, T., J. Hurley, J. Garcia, K. Yoder, A. Katz, D. Roberts, J. Cho, T. Kanigan, et al.** 2006. Nanoliter high throughput quantitative PCR. Nucleic Acids Res. *34*:e123.
32. **Bustin, S.A.** 2004. A-Z of Quantitative PCR. International University Line, La Jolla, CA.
33. **Pinkel, D., R. Segraves, D. Sudar, S. Clark, I. Poole, D. Kowbel, C. Collins, W.L. Kuo, et al.** 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. *2*:207–211.
34. **Amaratunga, D. and J. Cabrera.** 2004. Exploration and Analysis of DNA Microarray and Protein Array Data. John Wiley, New York.
35. **Amaratunga, D. and J. Cabrera.** 2009. A conditional t suite of tests for identifying differentially expressed genes in a DNA microarray experiment with little replication. Stat. Biopharmaceut. Res. *1*:26-38.
36. **Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter.** 1994. Molecular Biology of the Cell, 3rd ed. Garland Publishing, New York.
37. **Wouters, L., H.W. Gohlmann, L. Bijnens, S.U. Kass, G. Molenberghs, and P.J. Lewi.** 2003. Graphical exploration of gene expression data: a comparative study of three multivariate methods. Biometrics *59*:1131-1139.
38. **Su, A.I., M.P. Cooke, K.A. Ching, Y. Hakak, J.R. Walker, T. Wiltshire, A.P. Orth, R.G. Vega, et al.** 2002. Large-scale analysis of the human and mouse transcriptomes. Proc. Natl. Acad. Sci. USA *99*:4465-4470.
39. **Cleveland, W.S. and S.J. Devlin.** 1988. Locally-weighted regression: an approach to regression analysis by local fitting. J. Am. Stat. Assoc. *83*:596-610.

Received 2 January 2008; accepted 28 May 2008.

Address correspondence to Mariusz Lubomirski, Ph.D., Johnson & Johnson Pharmaceutical Research & Development, L.L.C., Welsh and McKean Roads, Spring House, PA, 19477, USA. e-mail: mlubomir@prdus.jnj.com