Quantitative Multiplexed Gene Expression: An Analysis of Sensitivity, Accuracy and Precision

Yong Wu, Mervin R. Gutierrez, Kathryn E. Sciabica, Jane Luo
Beckman Coulter, Inc.

Quantitative gene expression analysis is considered the gold standard by which all gene expression analysis tools are measured. The GenomeLab™ GeXP Genetic Analysis System from Beckman Coulter offers multiplexed, quantitative gene expression analysis capable of examining up to 30 genes in a single reaction from as little as 5 ng total RNA. Here, we demonstrate that the GeXP produces gene expression data with superb linearity and is sensitive enough to precisely detect even small changes in gene expression. A highly linear correlation between the amount of RNA and gene expression quantitation value was generated for each gene in a multiplex with an average correlation coefficient (R^2) well above 0.99. In addition, we verified that ten consecutive 0.5-fold increases in RNA concentration were accurately quantified by the GeXP for all 24 genes in the multiplexed assay. The capacity of GeXP to deliver multiplexed, sensitive and precise gene expression analysis opens a new door for scientists to explore subtle yet biologically meaningful changes in an effective and efficient manner.

Quantitative analysis of multiplexed gene expression in a single reaction, from a limited amount of total RNA, is of great interest to research scientists. Conventional techniques utilize either microarrays, which detect the expression of a high number of genes per reaction at high cost and low sample throughput, or real-time PCR* assays, which detect the expression of a few genes at moderate throughput. More often researchers have shifted their focus to 20 or more genes during biomarker validation, signal transduction or metabolic pathway studies. Following this shift, Beckman Coulter has recently developed the GenomeLab GeXP (Gene eXpression Profiler) Genetic Analysis System that can detect up to 30 genes in a single reaction containing 5 ng to 500 ng of total RNA. GeXP uses a patented universal priming strategy to overcome potential bias in amplified targets that are typically associated with other types of multiplexed assays. Since its inception, GeXP has demonstrated correlative results to real-time PCR and microarray data. (1-4 and unpublished data)

Equipped with eight capillaries and a two-plate format, the GeXP system is capable of analyzing 240 samples per 24-hour period. This produces 7,200 gene expression results for a 30-gene multiplexed assay. Researchers not only benefit from reduced labor and reagent expenditures, but also significant time savings due to expedited data collection with the GeXP system. Further, by providing more gene expression results per sample than real-time PCR, GeXP diminishes the experimental constraint imposed by limited quantities of RNA, such as those from formalin-fixed paraffin-embedded (FFPE) samples or tissue biopsies. The capabilities of GeXP has been increasingly recognized by academic and other research communities ranging from plant sciences and viral infection studies to cancer research.(1-4)

In this study, we demonstrate that GeXP has precise linearity in gene expression quantitation for each gene in a multiplex. The gene expression quantitation (GEQ) value, calculated from a standard curve, displayed highly linear correlation to the amount of RNA used. The average correlation coefficient (R^2) achieved was greater than 0.99. We further confirm that the GeXP delivers unique sensitivity in detecting subtle changes in RNA levels. The overall results indicate that the GeXP system is not only powerful enough to effectively deliver thousands of gene expression results per day, but also sensitive enough to detect small changes of gene expression levels in a very precise manner.

Materials and Methods

Reagents: GenomeLab GeXP Start Kit and GenomeLab GeXP Human Breast Cancer*Plex* Kit (Beckman Coulter, Inc.), ThermoStart** Taq DNA Polymerase (Thermo Scientific)

Reverse transcription, multiplex PCR and PCR fragment separation: cDNA synthesis from the reverse transcription of total RNA, multiplex PCR and PCR fragment separation on the GeXP system were performed following the standard GeXP Chemistry Protocol (P/N A29143) and GeXP Human Breast Cancer*Plex* Kit Insert (P/N A32698).

Optimization of the amount of KAN^r RNA in GeXP reactions: The data analysis method presented in this study utilized the signal level of a gene relative to the internal control, KAN^r gene. The original kanamycin resistance (KAN^r) RNA (5ng/uL) from the GenomeLab GeXP Start Kit was diluted 2-fold over 11 serial dilutions with RNA Storage Solution (Ambion). The final amount of KAN^r RNA in each reaction was in a range from 25 pg to 25 ng, and was added into a GeXP multiplex reverse transcription reaction containing 20 ng of Human Breast Cancer*Plex* (HuBC) control RNA from the Human Breast Cancer*Plex* Kit. The amount of KAN^r RNA that generated a peak height signal intensity equivalent to the median signal levels in the multiplex was selected. This optimal amount of KAN^r RNA (0.1 ng) was used in the reactions to establish the standard curves.

Establishment of standard curves: GeXP multiplex reactions were carried out using decreasing amounts (from 512 ng to 1 ng in 2-fold dilutions) of HuBC control RNA in the presence of the optimized amount of KAN^r RNA, 0.1 ng. For each amount of control RNA, six technical replicates were prepared for multiplex reverse transcription. Each multiplex reverse transcription reaction was independently used for a multiplex PCR reaction. Each multiplex PCR reaction was separated on a different capillary channel of a capillary array.

Determination of GeXP sensitivity in quantitation: A series of ten 0.5-fold increase of HuBC control RNA, from 6.5 ng to 250 ng were used in reactions containing 0.1 ng of KAN^r RNA. For each amount of control RNA, four technical replicates were prepared for multiplex reverse transcription. Each multiplex reverse transcription reaction was independently used for a multiplex PCR reaction. Each multiplex PCR reaction was separated on a different capillary channel of a capillary array.

GeXP fragment analysis: Fragments separated on the GeXP system were analyzed by a set of modified GeXP fragment analysis parameters (slope threshold = 1, peak height threshold = 0) in order to call all possible peaks including the smallest ones. Multiplex-specific fragments were selected by applying exclusion filters and were exported to eXpress Analysis software, where they were normalized against the KAN^r gene. Normalized data was exported to Microsoft Excel** for further analysis.

Data analysis: An XY (scatter) chart was used to plot the relationship between the relative signal level normalized against KAN^r (X-axis) and the amount of HuBC control RNA (Y-axis). A best fit regression line with a third-order polynomial was applied. The regression line was considered a standard curve and its equation was used to calculate the GEQ value from relative signal level normalized against KAN^r obtained from the eXpress Analysis report. The relative accuracy (RA) of a GEQ value is defined as: RA = (RNA amount – Absolute (RNA amount – GEQ value))/RNA amount x 100%.

Results

Establishment of the standard curve for a multiplex

Like many other quantitation methods, the quantitative analysis developed in this study uses a standard curve to calculate the GEQ value corresponding to a relative signal level. An overall standard curve was established to evaluate how precisely the average signal of a multiplex correlates to the amount of RNA. To establish the standard curve for Human Breast Cancer*Plex*, the fragment data from GeXP reactions performed with 2-fold serial dilutions of HuBC control RNA, ranging from 512 ng to 1 ng, were analyzed by eXpress Analysis software. Relative signal level of each gene was obtained by normalizing its peak area against that of the KAN^r gene. These relative signal levels were exported to Microsoft Excel from eXpress Profiler in the **Profile (by gene)** format. The average of the relative signal level of all 24 genes for each concentration of control RNA was then calculated.

A scatter chart was used to plot the relationship between the average of relative signal level of each multiplex reaction (x-axis) and the amount of HuBC control RNA (y-axis). A best fit regression curve with a third-order polynomial was applied (Figure 1). A coefficient of determination (R^2) of 0.9999 was obtained between the average relative signal level and the amount of HuBC control RNA. In regression analysis, the R^2 is a statistical measure of how well the regression line approximates the real data points. It represents the fraction of variability in the y-values that can be explained by the variability in the x-values. An R^2 of 1.0 indicates that the regression line perfectly fits the data. Therefore, a regression curve with high R^2 value, generally above 0.99, in GeXP data analysis can be considered a very reliable standard curve. The equation of the standard curve (regression curve) is used to calculate the gene expression quantitation (GEQ) value (y) from a relative signal level (x) obtained from the eXpress Analysis report. Therefore, a

GEQ value is a signal level quantitation derived from a standard curve.

The standard curve equation (Figure 1) for Human Breast Cancer*Plex* was used to calculate the GEQ value (y) from the relative signal level (x) for each technical replicate. When the average GEQ value from six technical replicates was plotted against the amount of HuBC control RNA, a superb linear correlation was achieved. The correlation coefficient (R^2 = 0.9998) between the amount of HuBC control RNA and GEQ value indicates a highly precise linear relationship between amount of RNA and the corresponding GEQ value calculated from the standard curve. (Figure 2).

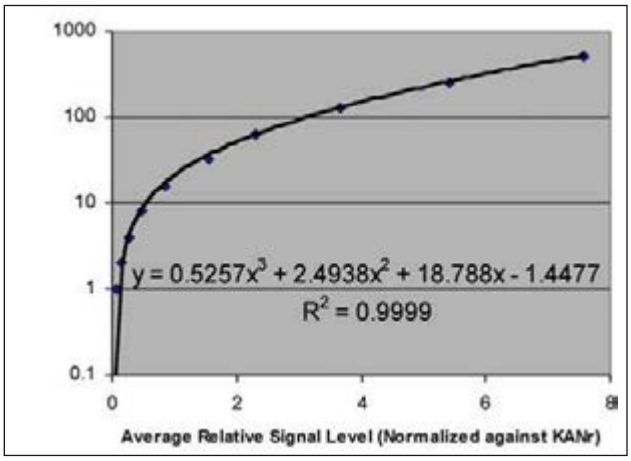


Figure 1. The correlation between the average of relative signal level for six technical replicates of all 24 genes in the Human Breast Cancer*Plex* and the amount of HuBC control RNA was near perfect in a third-order polynomial model. The fitting equation and R^2 are shown on the chart. The GEQ value (y) is calculated using the equation from this standard curve with relative signal level (x) for a particular RNA concentration.

To evaluate the accuracy of this data analysis method, the relative accuracy (RA) for each GEQ value was calculated using the equation described in the Materials and Methods. The average RA is 95.0% for the GEQ value corresponding to the amount of HuBC control RNA from 2 ng to 512 ng (Table 1). A smaller average RA value was observed when the amount of RNA was less than 4 ng, primarily due to lack of signal for some of the low expressers in the Human Breast Cancer*Plex*. The average RA increases to 97% if only data points from 4 ng to 512 ng of RNA are considered.

This approach also provides a high quality fitting model for each individual gene in the Human Breast Cancer*Plex*. By plotting the relative signal level of an individual gene against amount of HuBC control RNA, a trend line with a specific equation is obtained for each gene in the multiplex. This equation was then used to calculate the GEQ value for each gene. To briefly demonstrate how well this analysis works on individual genes, we selected three genes from the multiplex: a low expresser, a median expresser and a high expresser.

Gene GNAZ (NM_002073), a low expresser, shows a very strong correlation between the amount of HuBC control RNA and the GEQ value from 8 ng to 512 ng of HuBC control RNA (Figure 3a). A near perfect correlation coefficient (R^2) was also observed between amount of HuBC control RNA and the GEQ value within RNA range from 2 ng to 512 ng for a median expresser *bbc3* (U82987, Figure 3b) and a high expresser PRC1 (NM_003981, Figure 3c).

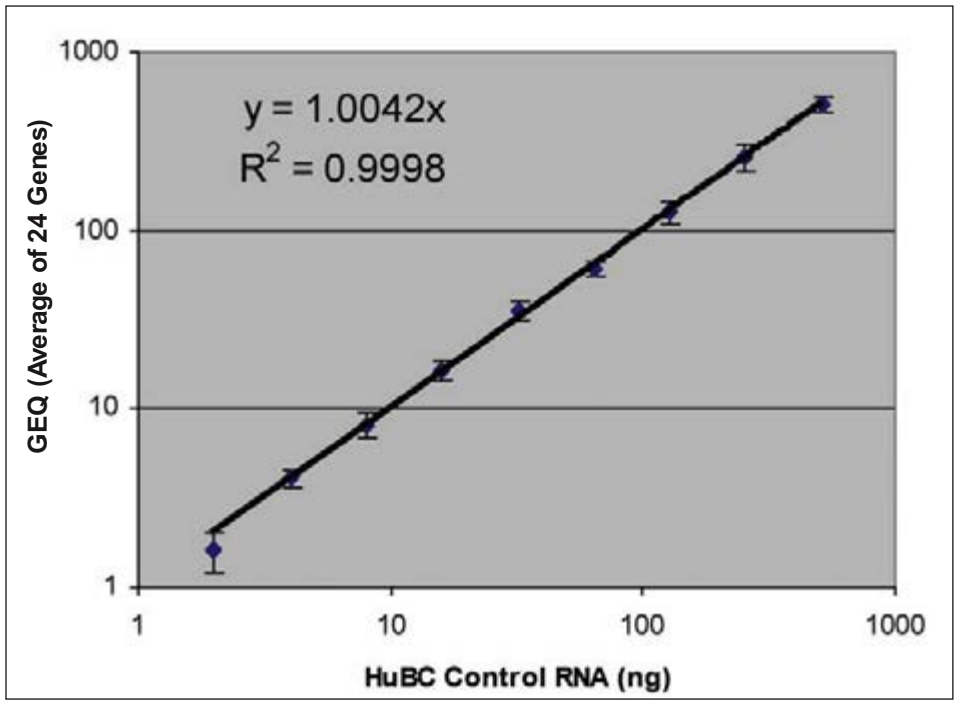


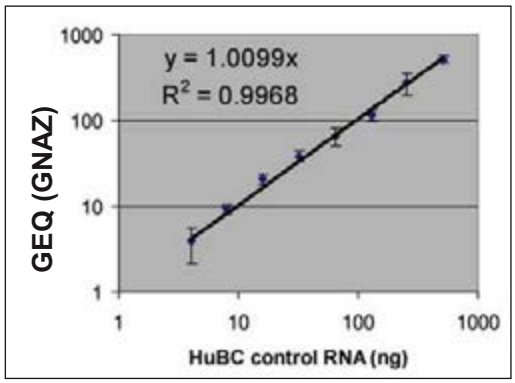
Figure 2. The linear correlation between the amount of HuBC control RNA and the GEQ value. The relationship between the amount of HuBC control RNA and the GEQ value from the average of relative signal level of all 24 genes in Human Breast Cancer*Plex* was plotted. A linear regression model was added to data points from 2 ng to 512 ng total input RNA. Error bars represent standard deviation of six technical replicates. The correlation coefficient (R^2) is displayed on the chart.

Table 1. Relative Accuracy for the GEQ value

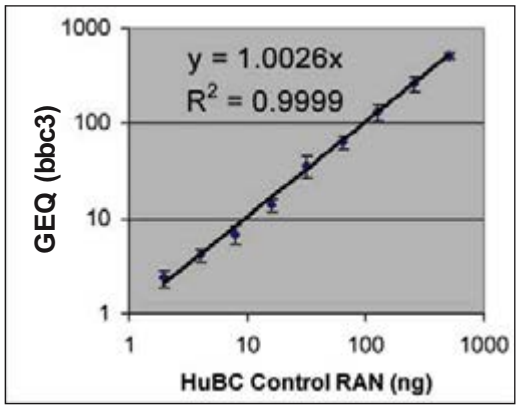
HuBC control RNA (ng)	2	4	8	16	32	64	128	256	512
GEQ value	1.6	4.1	8.2	16.5	35.5	61.4	126.5	257.6	511.6
RA (%)	79.4	98.2	97.7	97.0	89.0	95.9	98.8	99.4	99.9
Average RA (%)	95.0								

Figure 3. The linear correlation between the amount of HuBC control RNA and the GEQ value for gene GNAZ (a), bbc3 (b) and PCR1 (c). The relationship between the amount of RNA and the GEQ value for each gene was plotted with a linear trendline. Error bars represent one standard deviation from the mean for six replicates. The R^2 of each gene is displayed on the respective chart.

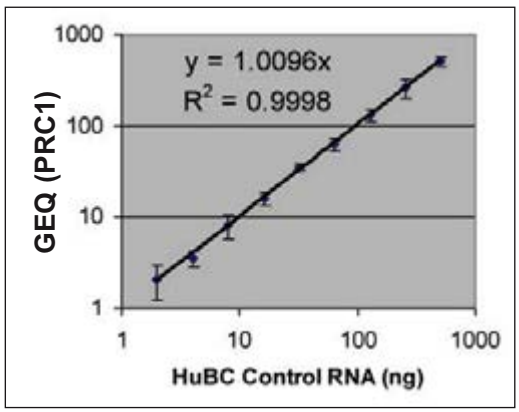
(a)



(b)



(c)



Determination of GeXP sensitivity in quantitation

To determine if GeXP is capable of detecting small changes in levels of gene expression, a 10-point dilution series varying the amount of HuBC control RNA in 0.5-fold increments, from 6.5 ng to 250 ng, was tested. A high correlation coefficient (R^2 = 0.9978) between the amount of RNA and the GEQ value was obtained (Figure 4). The %CV of this experiment, which reflects total variation in the GeXP process (reverse transcription, PCR and capillary electrophoresis), ranges from 6.6 to 16.5. The average RA of this assay is higher than 96% (Table 2). This result indicates that GeXP can confidently detect gene expression changes as small as 0.5-fold.

Each of the 24 genes in Human Breast Cancer*Plex* was analyzed with this method and the R^2 for each gene was obtained (Table 3). For all 24 genes, the average and median values for the R^2 were 0.9955 and 0.9976, respectively. These results demonstrate that GeXP delivers highly sensitive and precise quantitative gene expression analysis.

Discussion

Sensitivity, precision and accuracy are key aspects of quantitative gene expression analysis. In this study, our results demonstrate that the GenomeLab GeXP possesses high standards for all of these qualities in a multiplexed approach.

The sensitivity to detect minute yet biologically important changes in gene expression has been largely desired by scientists to study the impact of subtle gene regulation, particularly in research areas of cancer, neurosciences and obesity.[5-8] Currently, small changes in gene expression cannot be accurately measured by conventional methods. Therefore, the significance of these subtle changes during tumor progression and neurodegeneration is poorly understood. The sensitivity of GeXP brings gene expression analysis into exciting new territory where subtle yet biologically significant changes are quantified, leading to the discovery of finely tuned mechanisms of gene expression regulation.

Our results demonstrate that GeXP is more sensitive in detecting minute changes in levels of gene expression than real-time PCR. This is likely due to the much higher signal

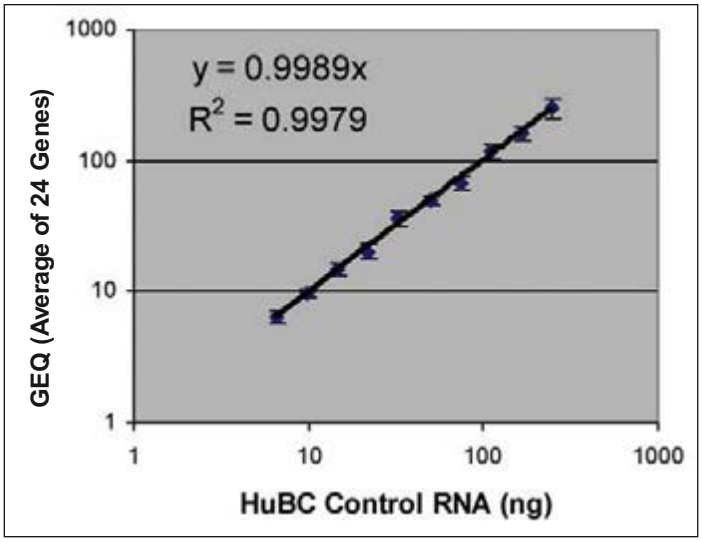


Figure 4. The linear correlation between the amount of HuBC control RNA and the GEQ value on detecting 0.5-fold change of target mRNAs. Error bars represent one standard deviation from the mean for four technical replicates. The correlation coefficient (R^2) is shown on the chart.

Table 2. Relative Accuracy (RA) of GeXP in detecting 0.5-fold change in amount of HuBC control RNA

HuBC control RNA (ng)	6.5	9.8	14.6	21.9	32.9	49.4	74	111	167	250
GEQ value	6.5	9.8	15.0	20.4	36.2	49.8	67.8	118.3	162.7	250.6
RA (%)	99.8	99.9	97.2	93.3	89.9	99.3	91.6	93.4	97.4	99.8
Average RA (%)	96.2									

Gene Symbol	Accession#	R^2
PPIA	BC000689	0.9984
GAPDH	NM_002046	0.9992
IGFBP5	L27560	0.9992
CP	NM_000096	0.9994
ESR1	NM_000125	0.9996
EXT1	NM_000127	0.9942
KRT18	NM_000224	0.9965
OXCT1	NM_000436	0.9985
DCK	NM_000788	0.9927
ACTB	NM_001101	0.9964
AP2B1	NM_001282	0.9945
GNAZ	NM_002073	0.9869
MYBL2	NM_002466	0.999
RFC4	NM_002916	0.9989
TGFB3	NM_003239	0.9993
CDC42BPA	NM_003607	0.9862
ALDH4A1	NM_003748	0.9977
WISP1	NM_003882	0.9811
PRC1	NM_003981	0.9963
KNTC2	NM_006101	0.9926
RAB6B	NM_016577	0.9913
EGLN1	NM_022051	0.9988
bbc3	U82987	0.9976
HSC4A2 IV2	X05610	0.9976
Average		0.9955
Median		0.9976

Table 3. For each gene in Human Breast Cancer*Plex* in detecting 0.5-fold changes

to noise ratio of GeXP. Both GeXP and real-time PCR have their lowest signal to noise ratio at the beginning of PCR and both generate increasingly stronger signal as the cycle number progresses. Real-time PCR generally analyzes data points at a signal level ten times higher than the standard deviation of background noise, known as the threshold cycle (C_t). In contrast, GeXP collects data points at a signal level a thousand to million times higher than background noise. Data collection at the lower signal to noise ratio may contribute to higher standard deviation for real-time PCR data. The user's guide for performing real-time PCR[9] specifies that the standard deviation of C_t value among technical replicates should not be greater than 0.3. A standard deviation of C_t value close to 0.3 reflects approximately 70% variation among technical replicates. In order for real-time PCR to be able to detect a 0.5-fold increase in mRNA level, the standard deviation of C_t value must be smaller than 0.1, which is very challenging under most circumstances.

A linear correlation between the GEQ value and the amount of RNA for each gene was achieved in GeXP multiplexed reaction with the average R^2 well above 0.99. Combined with the high signal to noise ratio, this linearity enables

GeXP to obtain a very high RA value of approximately 95%. In an evaluation study, real-time PCR demonstrated an RA value of 9.4, which is equivalent to an RA of 90.6% using the calculation method in this study.[10] The higher RA of GeXP over real-time PCR confirms that GeXP is not only geared for cost and time savings, but also delivers higher accuracy than conventional quantitative methods.
Another advantage of GeXP over one gene at a time assays is its accuracy in normalization. Because reference genes are analyzed in the same reaction with genes of interest, normalization in GeXP is not affected by pipette variation or well-to-well variation during the PCR process. Further, the ability of GeXP to include multiple housekeeping genes in one multiplex provides users the flexibility to choose a different housekeeping gene or use multiple housekeeping genes for a particular study.

Lower than average RA was found for the GEQ values corresponding to the amount of total RNA less than 4 ng for some low expresser genes in the multiplex. This can be attributed to the fact that when the amount of total RNA is low, there are too few copies of low expresser mRNA to generate reproducible signals. We recommend using a minimum of 5 ng of RNA for the multiplex reaction. This minimum amount remains much lower than the one required by real-time PCR and microarray assays. For median and high expressers, 2 ng of total RNA is usually enough to obtain reproducible results.

In summary, GeXP demonstrates superb linearity and accuracy in analyzing gene expression. Its ability to detect small but biologically significant changes distinguishes GenomeLab GeXP from other currently existing technologies. With the capacity of analyzing up to 30 genes in a single reaction, GeXP is equipped to deliver accurate and precise information in the most effective and efficient manner.

Acknowledgement

We greatly appreciate Jeff Chapman, Samira Kaissi, Sam Dougaparsad, Amber Volk and many others for reviewing the manuscript.

References

1. Chen QR, Vansant G, Oades K, Pickering M, Wei J, Song YK, Monforte J, and Khan J. **2007** Diagnosis of the small round blue cell tumors using multiplex polymerase chain reaction. *Journal of Molecular Diagnostics* 9: 80-88.

2. Boyer JD, Robinson TM, Kutzler MA, Vansant G, Hokey DA, Kumar S, et. al. **2007** Protection against simian/human immunodeficiency virus (SHIV) 89.6P in macaques after coimmunization with SHIV antigen and IL-15 plasmid. *Proceedings of the National Academy of Sciences* 104: 18648-18653.

3. Hayashi E, Aoyama N, Wu Y, Chi HC, Boyer SK, and Still DW. **2007** Multiplexed, quantitative gene expression analysis for lettuce seed germination on GenomeLab™ GeXP Genetic analysis System. *Application Information*, A-10295A, Beckman Coulter Inc. http://www.beckmancoulter.com/Literature/BioResearch/A-10295A.pdf

4. Kim B, Masuta C, Matsuura H, Takahashi H, and Inukai T. **2008** Veinal necrosis induced by *Turnip mosaic virus* infection in *Arabidopsis* is a form of defense response accompanying HR-like cell death. *Molecular Plant-Microbe Interactions* 21: 260-268

5. Koza RA, Nikonova L, Hogan J, Rim JS, Mendoza T, Faulk C, et. al. **2006** Changes in gene expression fore shadow diet-induced obesity in genetically identical mice. *PLoS Genetics* 2: 0769-0780

6. Yan H, Dobbie Z, Gruber SB, Markowitz S, Romans K, Giardiello FM, et. al. **2002** Small changes in expression affect predisposition to tumorigenesis. *Nature Genetics* 30: 25-26

7. Kaisera S, Foltza LA, Georgea CA, Kirkwoodb SC, Bemisc KG, Lind X, et. al. **2004** Phencyclidine-induced changes in rat cortical gene expression identified by microarray analysis: implications for schizophrenia. *Neurobiology of Disease* 16: 220-235

8. Bennett CN, Ross SE, Longo KA, Bajnok L, Hemati N, Johnson KW, et al. **2002** Regulation of Wnt signaling during adipogenesis. *Journal of Biological Chemistry* 277: 30998–31004.

9. Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR. Applied Biosystems.

10. Canales RD, Luo Y, Willey JC, Austermiller B, Barbacioru CC, Boysen C, et. al. **2006** Evaluation of DNA microarray results with quantitative gene expression platforms. *Nature Biotechnology* 24: 1115-1122.

For Research Use Only. Not for use in diagnostic procedures.

www.sciex.com/ce

View SCIEX products at www.sciex.com
Find your local office at www.sciex.com/offices

AB SCIEX Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
www.absciex.com