Application Note

Genomics

Agilent

Trusted Answers

Ultra-low Input Total RNA Gene Expression Analysis Using the Agilent SurePrint G3 Human Gene Expression Microarray

Authors

Makoto Ogino, PhD
Research Fellow, Drug Discovery Resrearch, Astellas Pharma, Inc.

Yayoi Fukuoka and Allant Kodaka, Agilent Technologies Japan, Ltd.

Abstract

Microarrays are a valuable tool for analyzing the expression level of multiple known transcripts and identifying transcriptomic profiles. Agilent provides high-quality microarrays, based on SurePrint technology, that enable analysis of total RNA extracted from different sample types, such as tissues and cultured cells. However, for some applications (e.g., tissue biopsy, rare samples, single cell), it is impossible to obtain the amount of total RNA required to perform the experiments, and a smaller amount of total RNA would generate lower signal intensity, which is insufficient for detection.

In this study, we modified the existing protocol to allow processing of samples starting from 250 pg of total RNA. The main changes implemented in the protocol are the use of a different cleanup kit that enables the purification of a smaller amount of labeled cRNA and increased hybridization time. The changes implemented to the original Agilent protocol allowed us to obtain valid signal intensity from as little as 250 pg of total RNA. We have demonstrated that this modified workflow can produce reliable data with excellent reproducibility and low background, enabling the analysis of a wide range of sample types that were previously impossible to analyze.

Introduction

Microarrays are a key tool for researchers interested in transcriptomic profiling. Microarrays are a simple and cost-effective method for analyzing global gene expression profiles of multiple samples, providing a balance of comprehensive data and fast sample-to-data processing time. Microarray-based expression profiling has been used in gene discovery, biomarker identification, disease classification, gene regulation, metabolic pathways, and much more. The Agilent platform is flexible, allowing users to work in single- and dual-color mode depending on the experimental question, and has a unique sensitivity that enables generation of reliable profiles with 5 logs of dynamic range, high reproducibility, and has shown good correlation with RNA-seq and RT-qPCR.

The standard protocol for the Agilent Low Input Quick Amp Labeling kit recommends using 10–200 ng of total RNA. This amount of RNA fits most applications; however in certain cases researchers need to analyze samples extracted from small tissues, biopsies, or rare specimens which can be as low as 1 ng. Other protocols are based on RNA amplification and labeling and allow users to start from limited starting material. However, amplification can introduce a bias that can be difficult to manage with data analysis.

This study demonstrates that with a few modifications to the original protocol, it is possible to detect transcripts starting with a lower amount of total RNA—as low as 250 pg—without any amplification steps.

Experimental

Two commercially available RNAs previously used during microarray quality control (MAQC) experiments were selected: Universal Human Reference RNA (Agilent p/n 740000, UR in this note) and Ambion First Choice Human Brain (Thermo Fisher Scientific p/n 6050, Br). Total RNA was quantified using the NanoDrop 1000 (Thermo Fisher Scientific) and quality was evaluated using the Agilent Bioanalyzer to determine purity and integrity. Sample information is summarized in Table 1.

Total RNA was diluted in order to obtain a titration curve during the experiments, yielding the final amounts of 10 ng, 2 ng, 1 ng, 500 pg, and 250 pg. Samples were diluted and brought to a final volume of 4.4 µl using nuclease-free water (Thermo Fisher Scientific, p/n 10977015).

Each starting amount was labeled in duplicate except for the 1 ng concentration, which was labeled in quadruplicate. Diluted spike-in (from spike-in mix) and T7 promoter primer (from the Low Input Quick Amp Labeling kit) were mixed as described in Table 2, then 0.9 µl of the mixture was added to each tube of total RNA, mixed by pipetting, then spun down. Because the pipetting volumes are very low for individual reactions in this modified protocol, we recommend making a master mix of at least 8 samples with overage as noted in Table 2.

Table 1. Sample provider and quality information.

Sample	Provider	P/N	Concentration	RIN
Universal Human Reference RNA	Agilent	740000	56.8 pg/µl, 113.6 pg/µl, 227.3 pg/µl, 454.5 pg/µl and 2.27 ng/µl	9.2
Ambion First Choice Human Brain	Thermo Fisher Scientific	6050	56.8 pg/ µl, 113.6 pg/µl, 227.3 pg/µl, 454.5 pg/µl and 2.27 ng/µl	8.4

Table 2. Master mix for spike in and T7 promoter primer.

	1 reaction	8 reaction (with excess)
Diluted spike-in	0.1 µl	1.0 µl
T7 promoter primer	0.8 µl	8.0 µl
Total	0.9 µl	9.0 µl

Table 3. Modified steps from original protocol.

	Original protocol	Modified method
Amount of total RNA	10-200 ng	0.25, 0.5, 1, 2 and 10 ng
Volume for labeling	1.5 µl/sample	4.4 µl/sample
Cleaning up of labeled cRNA	QIAGEN RNeasy Mini Kit (elution: 30 µl)	QIAGEN RNeasy Micro Kit (elution: 14 µl)
cRNA amount for hybridization	600 ng	600 ng or all of available cRNA
Hybridization	17 hours	41 hours

The Agilent spike-in controls are diluted for this modified protocol in a similar manner as the standard protocol, with the exception that the final dilution is 20-fold more concentrated, because 20-fold less volume of spike-in is used in the final reaction. For example, in the standard protocol the Agilent spike-in controls were diluted 1:100,000 and 2 µl was used for a starting amount of 10 ng total RNA. In this new protocol the Agilent spike-in controls are diluted 1:5,000 and 0.1 µl is used for a starting amount of 10 ng total RNA.

After annealing the T7 promoter primer to the total RNA and spike-in controls, the samples were incubated to perform reverse transcription and labeling with Cy3-CTP following the original protocol (One-Color Microarray-Based Gene Expression Analysis ver.6.9, G4140-90040).

The cleanup protocol of labeled cRNA was modified to concentrate the final labeled cRNA sample. The QIAGEN RNeasy Micro Kit was used instead of the RNeasy Mini kit suggested by the original protocol. The QIAGEN RNeasy Micro Kit was optimized to deliver consistent RNA yields from very small amounts of starting material and to elute them in a smaller amount of buffer, enabling sample concentration for downstream applications. The modified steps are summarized in Table 3.

After labeling and purification, the concentration of labeled cRNA was measured using the NanoDrop 1000 in microarray measurement mode. Labeled cRNA were hybridized to SurePrint G3 Human Gene Expression microarray 8x60K ver.3 at 65 °C for 41 hours instead of 17 hours.

Extending the hybridization time is key for allowing hybridization kinetics to reach completion and ensure hybridization on the microarray surface of all the labeled molecules present in solution, thus increasing signal to enable analysis of low-concentration samples.

After hybridization, the array slides were washed following the original protocol and scanned using the Agilent SureScan Microarray Scanner. Scanned data were analyzed using both Agilent Feature Extraction and Agilent GeneSpring GX software. When analyzing data across different microarrays, Agilent recommends performing a 75th percentile shift of the signals on each microarray before comparison.

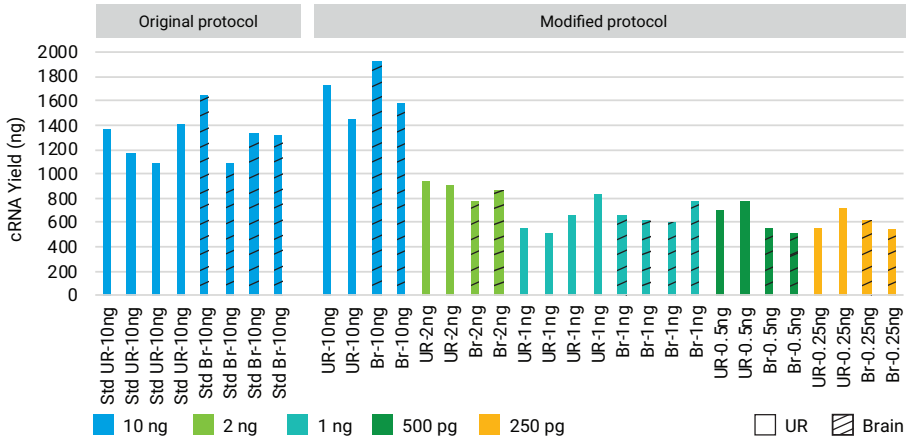


Figure 1. A comparison of the yield of labeled cRNA between the original and modified protocol.

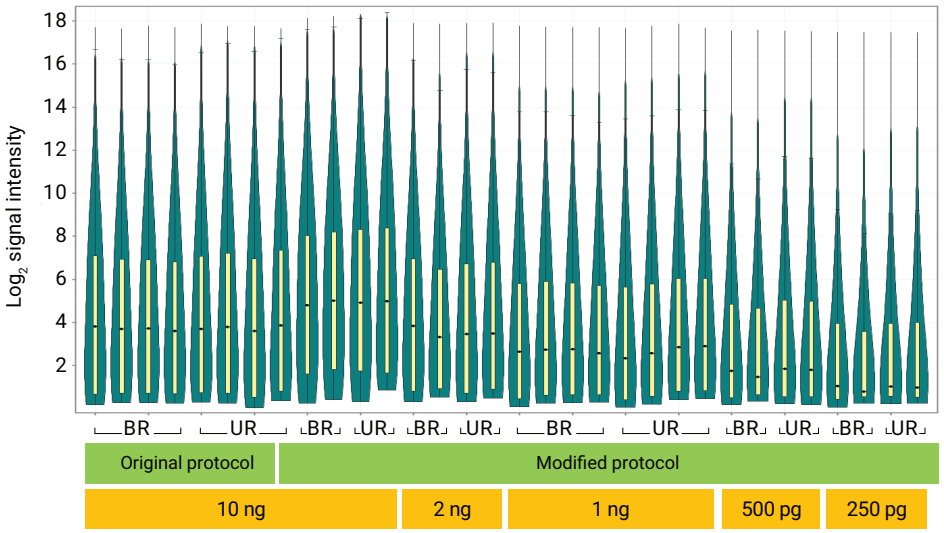


Figure 2. Distribution of the entire signal intensity from each microarray. Signal distribution in the image is represented by the yellow line. The lower base is the 25th percentile, the upper base is the 75th percentile, and the black dots represent median signal intensity for each array.

Results and discussion

Yield of labeled cRNA

Labeled cRNA yield depends on the starting amount of total RNA used for labeling (Figure 1). For half of the samples, starting amounts equal to or lower than 1 ng failed to produce the recommended 600 ng yield required for hybridization of an 8-pack array. In addition, it was impossible to calculate specific activity for low-yield samples, since the absorbance of Cy3 (A550) was below the detection limit of the NanoDrop 1000.

Despite this result, we decided to proceed with the protocol. Fragmentation was performed following the original Agilent protocol, and samples were hybridized on an array. When possible, each array was loaded with 600 ng of labeled cRNA. When the required amount of 600 ng was not reached, we proceeded using the entire volume of RNA obtained with labeling. Due to the reduced elution volume in the QIAGEN RNeasy Micro Kit, samples are eluted in 14 µl of nuclease-free water, enabling the addition of the entire volume to the hybridization mix without any concentration step.

Quality of array data

As expected, data obtained from 10 ng samples using the original and modified protocols show differences in signal intensity. Signal intensity obtained with the modified protocol is higher than with the original protocol. This is due to longer hybridization time, allowing hybridization kinetics to reach completion (Figure 2).

Net signal statistics are an indication of the dynamic range of the signal on a microarray for both non-control probes and spike-in probes.

The QC report uses the range from the 1st to the 99th percentile as an indicator of dynamic range for the microarray. When analyzing low-input samples labeled with the modified protocol, the signal intensity level decreased according to the starting amount, as expected. Encouragingly, the dynamic range remained stable for all the samples, showing some loss of signal for the very low-expressed genes in the 0.5 and 0.25 ng samples. This data highlights the robustness, sensitivity, and specificity of the Agilent Gene Expression platform.

The Agilent Feature Extraction software extracts data from images and generates a three-page pdf QC report for each array, including statistical results to help evaluate the reproducibility and reliability of each individual microarray. Among the different parameters generated, the *gNonCntrlMedCVProcSignal* values represent the coefficient of variation (CV%) of signal intensity between replicated probes. A lower median CV% value indicates better reproducibility of signal across the microarray across the microarray, ideally the CV% value should be 8% or less. In this study, low-input samples showed slightly increased CV% values compared to higher-input samples. However, most of the samples have a CV% lower than 8%, which is still in a suitable range for downstream analysis (Figure 3).

Negative control stats give a rough estimate of the background noise on the microarray. The metric *gNegCtrlAveNetSig*, which indicates the background of each array, is very stable between arrays (Figure 4). All *gNegCtrlAveNetSig* values were well under the upper limit of 40 counts, and they did not increase even though hybridization was prolonged up to 41 hours. This clearly indicates that increased hybridization time allows all molecules in the solution to hybridize, without affecting stringency and specificity. Higher background would be detrimental for detecting low-expressed transcripts, especially when working with low-input RNA; therefore, it is critical to maintain low background values, as demonstrated here.

Another statistical parameter linked to dynamic range is *gNonCtrlNumWellAboveBG*, a measure of the number of non-control features whose signals are well above background. Used as a metric for the number of features with significant signal, this parameter is affected by background and signal intensity from

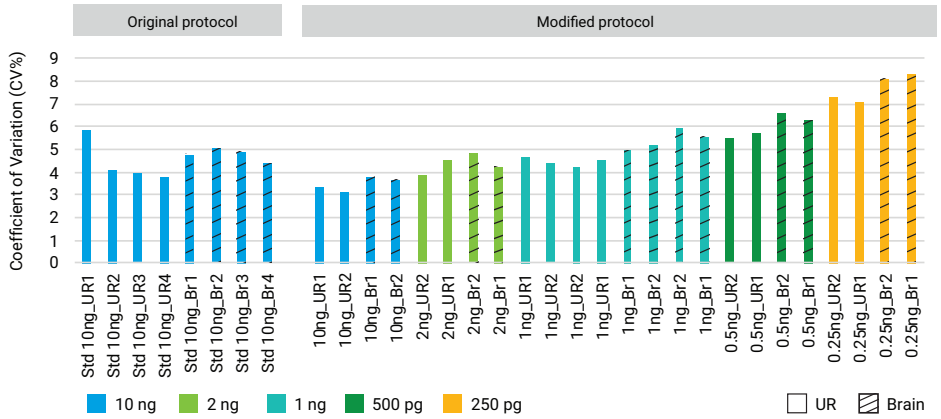


Figure 3. Reproducibility of statistical data as shown by coefficients of variation (CV%) of replicated probes. CV% increases when sample input is reduced; all samples are below the 8% CV recommended by the protocol.

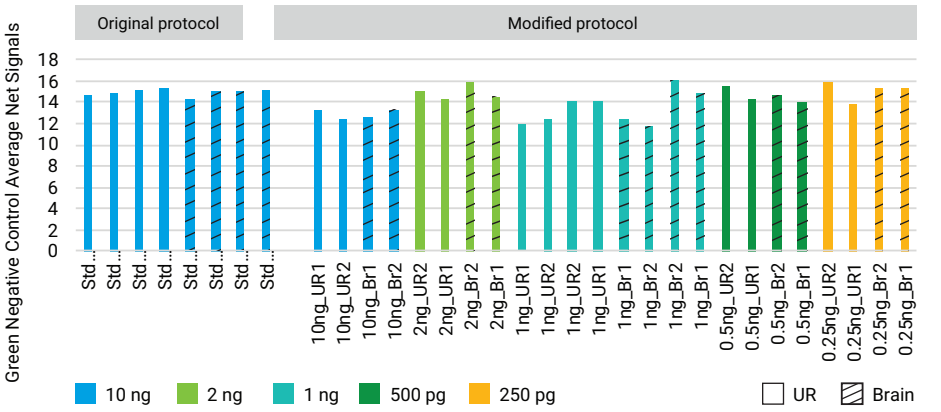


Figure 4. Signal from negative control probes (*gNegCtrlAveNetSig*). Negative control probes are designed to form a loop during hybridization, which prevents any hybridization. Their signal is an evaluation of background on the array.

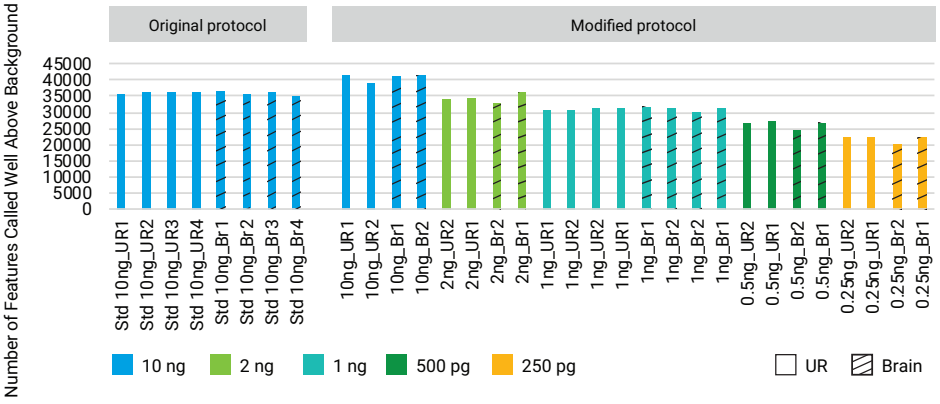


Figure 5. Measure of the number of biological non-control features whose signals are well above background. This is used as a metric for the number of features with significant signal.

each sample. As expected, low-input RNA samples have a lower amount of detected probes, mainly because of the inability to detect the very low-expressed genes (Figure 5).

Reproducibility

Measurement of reproducibility is key for defining robustness and specificity of the platform. In order to evaluate the impact on experimental results of total RNA starting amount and changes implemented in the protocol, we compared signals obtained at different concentrations. The correlations, or R^2, were sightly inferior when the 10 ng input samples were compared to lower RNA input; still, all correlations were well above $R^2 > 0.9$.

The Agilent SurePrint G3 Human Gene Expression microarray 8x60K ver.3 includes probes to long noncoding RNA (lncRNA) based on LNCipedia and Broad Institute in addition to mRNA. Figure 6 shows detected probes in each pair of arrays, with the points corresponding to lincRNA expression colored light blue. The dynamic range of mRNA and lncRNA probes was over 4 logs, even starting with as low as 250 pg of RNA.

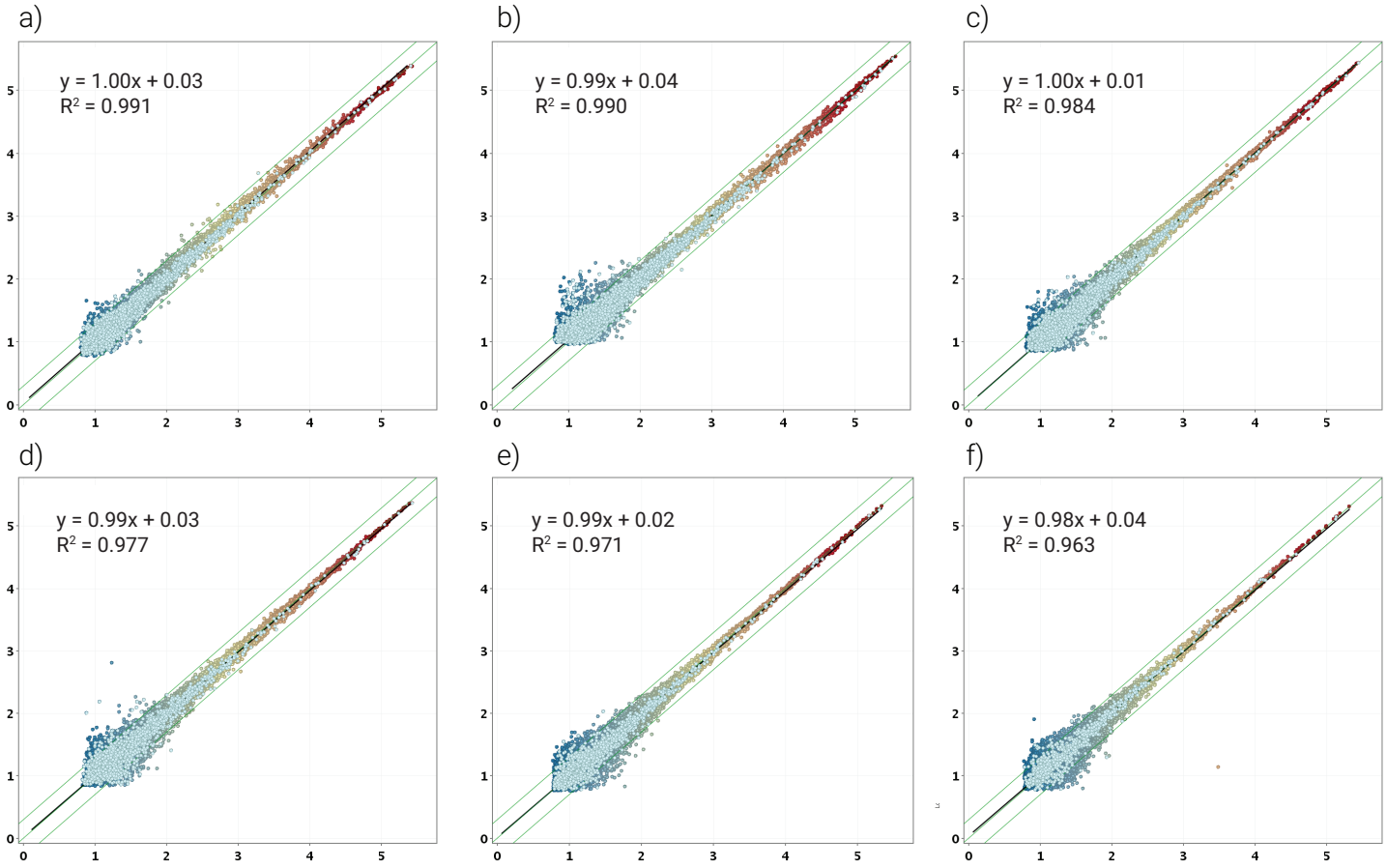


Figure 6. Scatter plot comparing the signals for biological probes for each starting amount of Universal Human Reference RNA. Light blue dots represent signals for long noncoding (lnc) RNAs. Log_{10} signal intensity without normalization between arrays. **a)** 10 ng labeled with the original protocol; **b)** 10 ng labeled with the modified protocol; **c)** 2 ng labeled with the modified protocol; **d)** 1 ng labeled with the modified protocol; **e)** 500 pg labeled with the modified protocol; **f)** 250 pg labeled with the modified protocol.

Differentially expressed genes

Figure 7a shows a scatter plot comparing $\log_2$ ratios for the 10 ng starting amount using the two different protocols described in this note. The correlation of $\log_2$ gene expression ratios in this plot was very tight, with R^2 of 0.965, indicating that modifications in the protocol are not impacting performance. Excellent reproducibility in gene expression ratios is also evident when the new modified protocol data with 10 ng input (on the x-axis) is compared to the reduced input amounts (on y-axis, Figure 7 b-e). The high concordance of gene expression ratios between very low RNA input amounts using the modified protocol as compared to the standard protocol with 10 ng input in Figure 7 indicate that the microarray assay generates gene expression log ratios that are linear down to very low gene expression levels.

The number of transcripts commonly called as up- or down-regulated with the different input RNA levels are shown in Figure 8. There were 419 transcripts that were commonly called as up and 492 transcripts were commonly called as down using GeneSpring. Additionally the number of transcripts called as up- or down-regulated depends on the starting amount of total RNA. This is an important parameter, as it clearly indicates that when running a study, it is better to define a starting amount and label the same amount of total RNA for all samples that need to be compared or analyzed together.

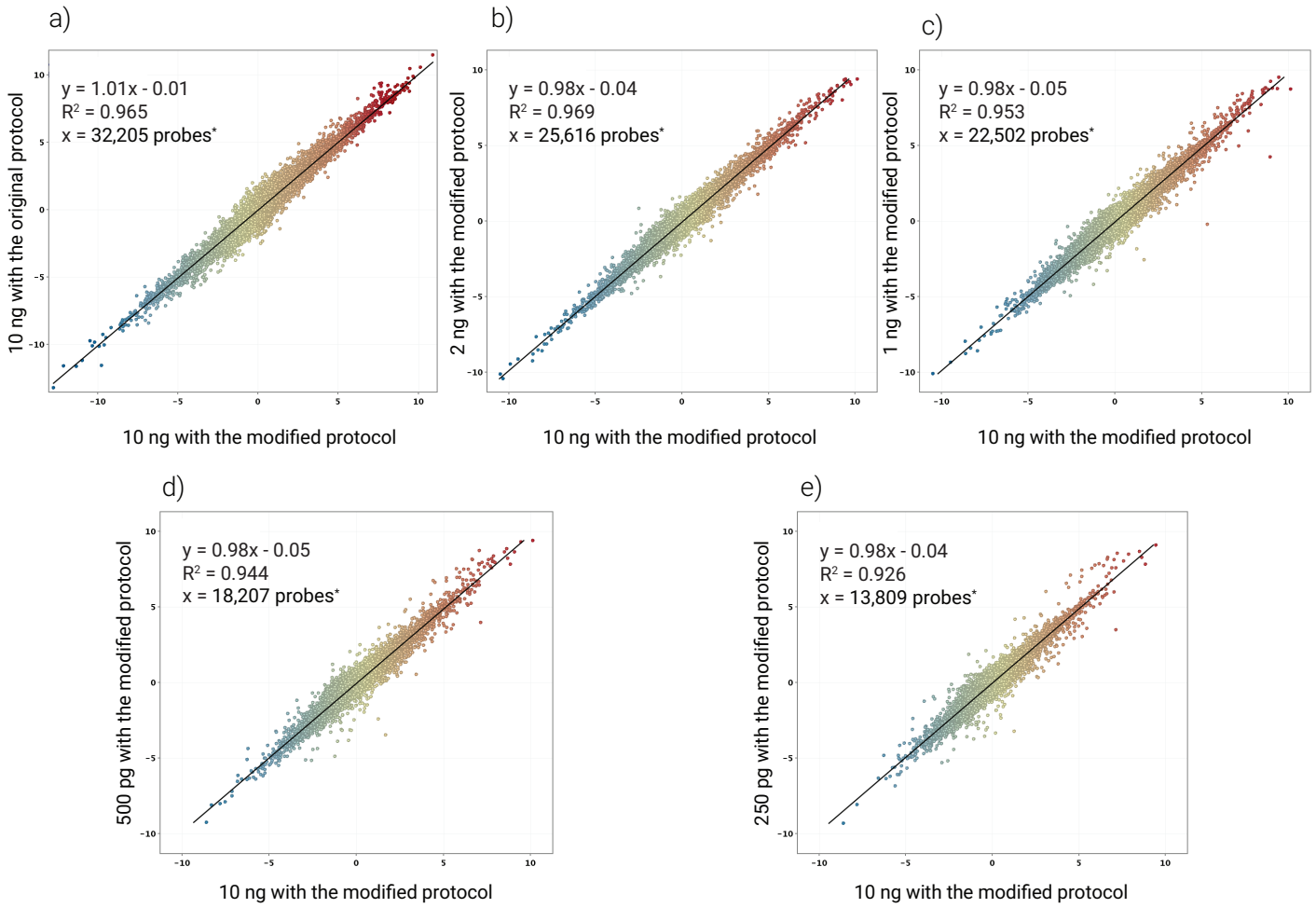
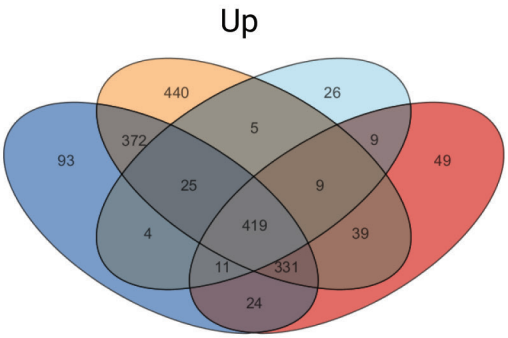
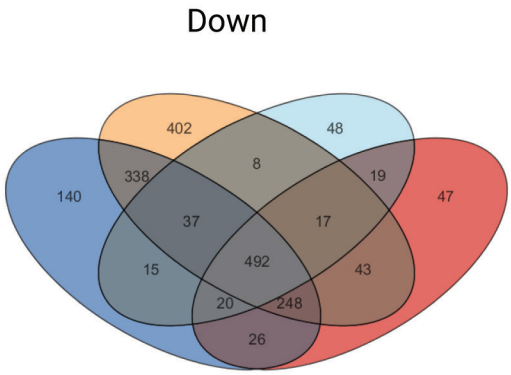


Figure 7. A comparison of $\log_2$ ratio (Brain/UR) between 10 ng with the modified protocol (x- axis) and the difference starting amounts.
* Number of probes called as 'Detected' in all 4 arrays.



Color	Experiment Name	Count in Experiment
	2 ng 75pcnt BvsU	1640
	1 ng 75pcnt BvsU	1279
	05 ng 75pcnt BvsU	891
	025 ng 75pcnt BvsU	508



Color	Experiment Name	Count in Experiment
	2 ng 75pcnt BvsU	1585
	1 ng 75pcnt BvsU	1316
	05 ng 75pcnt BvsU	912
	025 ng 75pcnt BvsU	656

Figure 8. Venn diagram showing the number of genes detected as up- and down-regulated by different experiments performed with variable starting amount of RNA. It is interesting to note that the number of genes detected in common is directly proportional to sample RNA input. (Corrected p-value cut-off: 0.05; fold change cut-off: 5.0; T Test: unpaired.)

Conclusion

There are many types of samples from which only a small amount of total RNA might be extracted. This study demonstrates that the amount of total RNA can be reduced to 250 pg for the Agilent SurePrint G3 microarray 8x60K by modifying a few steps from the original workflow. These modifications allow users to avoid additional amplification that might create bias. Our results returned higher signal intensity with longer hybridization, without an increase in background, allowing detection of low-expressed transcripts. The smallest amount tested in this study was 250 pg, but less total RNA might work if the purpose is to detect high-expressed transcripts.

Non Agilent products used in this note.

Products	Manufacturer
Ambion First Choice Human Brain	Thermo Fisher Scientific
UltraPure™ DNase/RNase-Free Distilled Water	Thermo Fisher Scientific
RNeasy Micro kit	QIAGEN
1.5 ml RNase-Free Tubes	Thermo Fisher Scientific
NanoDrop 1000	Thermo Fisher Scientific
Thermomixer comfort	Eppendorf
Water bath	TOKYO RIKAKIKAI
microcentrifuge	Eppendorf
Dish and rack	Thermo Fisher Scientific
Magnetic stirrer	IKEDA SCIENTIFIC Co.LTD / AS ONE

www.agilent.com

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

This information is subject to change without notice.

PR7000-2337

Printed in the USA, November 6, 2019
5994-1512EN

Agilent
Trusted Answers