BIOINFORMATICS ORIGINAL PAPER

Vol. 27 no. 9 2011, pages 1284–1289
doi:10.1093/bioinformatics/btr135

Gene expression

Advance Access publication March 16, 2011

A wholly defined Agilent microarray spike-in dataset

Qianqian Zhu[1,2,†], Jeffrey C. Miecznikowski[2,3], and Marc S. Halfon[1,4,5,6,*]

[1]Department of Biochemistry, [2]Department of Biostatistics, State University of New York at Buffalo, Buffalo, NY 14214, [3]Department of Biostatistics, Roswell Park Cancer Institute, Buffalo, NY 14263, [4]New York State Center of Excellence in Bioinformatics and Life Sciences, Buffalo, NY 14203, [5]Department of Biological Sciences, State University of New York at Buffalo, Buffalo, NY 14260 and [6]Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY 14263, USA

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Spike-in datasets provide a valuable resource for assessing and comparing among competing microarray analysis strategies. Our previous wholly defined spike-in datasets, the Golden and Platinum Spikes, have provided insights for the analysis of Affymetrix GeneChips. However, a similar dataset, in which all cRNA identities and relative levels are known prospectively, has not been available for two-color platforms.

Results: We have generated a wholly defined spike-in dataset for Agilent microarrays consisting of 12 arrays with more than 2000 differentially expressed, and approximately 3600 background, cRNAs. The composition of this 'Ag Spike' dataset is identical to that of our previous Platinum Spike dataset and therefore allows direct cross-platform comparison. We demonstrate here the utility of the Ag Spike dataset for evaluating different analysis methods designed for two-color arrays. Comparison between the Ag Spike and Platinum Spike studies shows high agreement between results obtained using the Affymetrix and Agilent platforms.

Availability: The Ag Spike raw data can be accessed at http://www.ccr.buffalo.edu/halfon/spike/index.html and through NCBI's Gene Expression Omnibus (GEO; accession GSE24866).

Contact: qzhu@buffalo.edu; mshalfon@buffalo.edu.

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on October 27, 2010; revised on February 11, 2011; accepted on March 9, 2011

1 INTRODUCTION

Since their introduction in the mid-1990s (Lockhart *et al*., 1996; Schena *et al*., 1995), DNA microarrays have become a leading tool in a diverse assortment of research disciplines, including cancer and other disease research, developmental biology and evolutionary biology. The popularity of microarrays owes much to their ability to simultaneously interrogate the expression level of tens of thousands of genes, enabling genome-scale transcriptional profiling of multiple tissues in essentially a single experiment. Although next-generation sequencing approaches such as RNA-Seq provide potentially compelling advantages over microarray-based gene expression profiling—no limitation to known transcripts, no cross-hybridization issues and a broader detection range for expression levels (Wang *et al*., 2009)—comparisons between these two technologies have indicated that RNA-Seq may have less accurate measurement of absolute RNA expression level (Willenbrock *et al*., 2009), less sensitive detection of differentially expressed genes (DEGs) (Git *et al*., 2010) and a higher error rate for weakly expressed genes (Liu *et al*., 2011). For the present, therefore, microarrays remain an important, cost-effective and mature platform for gene expression profiling.

The benefit of profiling numerous genes simultaneously by microarrays leads to one of the technique's biggest drawbacks: it is generally not possible to independently verify such a large number of individual data points, making it difficult to assess the accuracy of the data analysis. This is of particular concern due to the tremendous number of microarray data analysis methods, which have proliferated in tandem with the increased use of the arrays themselves.

Most previous studies of microarray performance, including the microarray quality control (MAQC) project, used RNA samples in which the identities of most of the individual RNA transcripts were not known (Affymetrix Latin square data, www.affymetrix.com/support/technical/sample_data/datasets.affx; Irizarry *et al*., 2003; Kerr, 2007; McCall and Irizarry, 2008; Patterson *et al*., 2006; Zahurak *et al*., 2007). Such an experimental design limits the evaluation of accuracy to only a small number of spike-in cRNAs present in the sample. We previously generated two wholly defined control datasets for the Affymetrix GeneChip platform, the Golden Spike and Platinum Spike datasets (Choe *et al*., 2005; Zhu *et al*., 2010). In these 'spike in' experiments, the identities and relative concentrations of the complete complement of present cRNAs are known. Over 1300 cRNAs are differentially expressed with designated fold changes ranging from 1.2 to 4, and over 2500 cRNAs serve as background with constant expression levels. We and others have used these datasets to determine optimal methods for analysis of Affymetrix microarrays under various experimental conditions (Chen *et al*., 2007; Choe *et al*., 2005; Hochreiter *et al*., 2006; Pearson, 2008; Schuster *et al*., 2007a, b; Turro *et al*., 2007; Zhu *et al*., 2010). Unfortunately, similar wholly defined spike-in datasets do not exist for other popular microarray platforms, in particular for dual-channel platforms. Different from Affymetrix's single-channel design, for dual-channel microarrays the two samples to be compared are labeled separately with different fluorescent dyes [e.g. cyanine-3 (Cy3) and cyanine-5 (Cy5)] and

*To whom correspondence should be addressed.
†Present address: Center for Human Genome Variation, Duke University, Durham, NC 27708, USA.

then hybridized simultaneously to the same array, allowing for direct comparison of differential expression. These differences in array technology require differences in data analysis. For instance, Affymetrix arrays typically call for interpreting both perfect match (PM) and mismatch (MM) probe values and combining the separate probe pairs into a probe set level expression summary (although some Affymetrix platforms, such as whole-transcript expression arrays, use only PM probes). Two-color arrays, on the other hand, require normalizing between fluorophore channels, correcting for possible differential incorporation of the fluorescent dyes, and comparing intensity ratios rather than absolute intensities to detect DEGs (Churchill, 2002; Duggan *et al*., 1999).

In order to facilitate evaluation of the various analysis methods required for dual-channel arrays, we generated a fully controlled spike-in dataset for one such platform, Agilent. We refer to this dataset as the 'Ag Spike' experiment due to its use of the Agilent platform and to denote its relationship to the Golden (Au) and Platinum (Pt) Spike studies. The Ag Spike dataset has a cRNA composition identical to that in our previous Platinum Spike dataset and consists of 12 arrays including technical replicates and dye swapped arrays. To illustrate the utility of the Ag Spike dataset, we evaluated 26 analysis routes, derived from combining several methods in individual steps of the analysis procedure, for their performance in detecting DEGs. In addition, we compared the results from this dataset and the Affymetrix-based Platinum Spike dataset and found high concordance across platforms. The Ag Spike dataset comprises a useful resource for further studies of optimal analysis methods for dual-channel microarrays.

2 METHODS

2.1 cRNA sample preparation and hybridization

We used the same 28 PCR pools that had been used before to generate the Platinum Spike dataset. These pools contain PCR products from 5725 *Drosophila* Gene Collection release 1.0 (DGCr1) cDNA clones. We mixed PCR products from all pools together with specified relative abundance to generate two samples representing the A and B conditions. Both samples were *in vitro* transcribed and then labeled with Cy3 and Cy5 fluorescent dye. Labeled A samples were hybridized with B samples labeled with reverse dye to three Agilent *Drosophila* Gene Expression microarrays. The process from *in vitro* transcription to hybridization was repeated twice (Fig. 1). Note that because samples were mixed as a pool, cRNAs from the same pool have identical fold change values between conditions (Table 1). At the end, we generated 12 arrays that were randomly placed in three independent slides to remove potential systematic error due to array position (Supplementary Material).

2.2 Probe fold change assignment

We first identified the probes hybridized to cRNAs present in samples by aligning the sequences of all probes to the obtained clone sequences. Alignment was done using BLAST (Altschul *et al*., 1990) (version 2.2.18) with word size seven, no complexity filter and *e*-value cutoff 1. In order to assign a probe to a specific clone, we required no fewer than 40% of the probe sequence matched identically to the clone sequence on the correct strand with no gaps (Supplementary Material). Probes assigned to multiple clones or to clones with multiple fold change values were excluded from our analysis. Therefore, using relatively low sequence identity cutoff in clone assignment step allowed us to remove potential cross-hybridization. For each probe uniquely assigned to a clone, the fold change value of the corresponding clone was assigned to the probe. If this clone had multiple

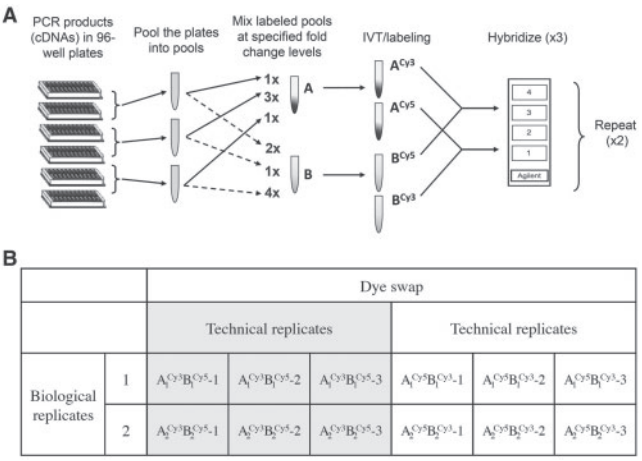

Fig. 1. Design and structure of the Ag Spike experiment. (**A**) Design of the Ag Spike experiment. PCR products were collected into 28 distinct pools, and then mixed at specified amounts to samples A and B, representing two different conditions. Each sample was *in vitro* transcribed and labeled with both Cy3 and Cy5 dyes. Differently labeled A and B samples were hybridized together in triplicate. The process from *in vitro* transcription to hybridization was repeated twice (Section 2). (**B**) Structure of the 12 Ag Spike arrays. This schema does not reflect the actual array position on slides.

Table 1. Number of DGCr1 clones and assigned fold change for each PCR pool

Pool name	Number of clones	Number of assigned probes[a]	Relative amount in A	Relative amount in B	Designated fold change (A versus B)
1	170	306	1	1.2	0.83
2	192	370	2	1	2.00
3	192	345	1.5	1	1.50
4	192	353	1	2.5	0.40
5	187	417	1	1	1.00
6	116	219	3	1	3.00
7	192	343	3.5	1	3.50
8	192	338	1	1.5	0.67
9	192	368	1	4	0.25
10	192	297	1.7	1	1.70
11a	192	402	1	1	1.00
12a	121	250	1	1	1.00
13a	192	319	1	1	1.00
14a	192	323	1	1	1.00
15a	192	370	1	1	1.00
16a	191	396	1	1	1.00
17a	139	267	1	1	1.00
18a	192	337	1	1	1.00
19a	192	334	1	3.5	0.29
11b	191	417	1	1	1.00
12b	223	417	1	1	1.00
13b	237	415	1	1	1.00
14b	288	575	1	1	1.00
15b	288	618	1	1	1.00
16b	288	619	1	1	1.00
17b	288	548	1	1	1.00
18b	250	449	1	1	1.00
19b	265	399	3.5	1	3.50

[a]There are 702 probes assigned to clones present in different pools.

potential sequences, we required that no fewer than 90% of the potential sequences mapped to the probe in order to assign the fold change value to the particular probe (Supplementary Material).

2.3 DEG detection

2.3.1 Preprocessing Background correction was either skipped or performed using one of the two different methods: *normexp* and *rma*. Within-array normalization was either not performed or performed with *loess* normalization. Later on between-array normalization step was skipped or performed using *quantile* normalization. We have previously found that normalization among all arrays performed the best when up- and down-regulation was balanced between the two compared conditions (Zhu *et al*., 2010). Therefore all 12 arrays in Ag Spike dataset were normalized together during the between-array normalization step. These preprocessing methods were from the *LIMMA* Bioconductor library (Ritchie *et al*., 2007; Smyth and Speed, 2003). We also performed preprocessing using only Agilent's *Feature Extraction* software (FE; version 9.5.3.1), followed by testing for DEGs as described below.

2.3.2 DEG tests Two different methods were used for prediction of DEGs: *fold change* (simply calculating $\log_2$ fold change between the A and B conditions) or *LIMMA* (Smyth, 2004). In this step, we treated technical replicates as biological replicates, because the variation introduced by independent sample preparation was much smaller than the variation introduced by hybridization (Supplementary Fig. S1).

We did note that the probe intensities from one array ($A_1^{Cy5}B_1^{Cy3}$-2) are higher than those of the other arrays, including its own corresponding technical replicates (Supplementary Fig. S2). This suggests that a technical error occurred in the handling of this array. However, the preprocessing procedures, including background correction and normalization, were able to bring this array comparable with other arrays [Pearson's correlation coefficient between this array and its technical replicates improves from 0.420 and 0.530 to 0.842 and 0.845 (Supplementary Fig. S3), and the median correlation between this array and its biological replicates is 0.758 after preprocessing], and the best performance for DEG detection using all 12 arrays was found to be slightly better than when excluding the particular array (data not shown). Therefore, our analysis included all 12 arrays in the Ag Spike dataset.

2.4 Receiver operating characteristic curves

For routes using *fold change* to detect DEGs, the absolute values of $\log_2$ fold change were given to the ROCR library (Sing *et al*., 2005) to generate the receiver operating characteristic (ROC) curves and calculate the area under the curve (AUC) values. For routes using *LIMMA*, the AUC values were calculated based on the absolute values of the corresponding test statistics.

2.5 Variance component analysis

We fitted the raw foreground intensities corresponding to a given probe across all arrays from the same condition with a mixed effect model: $\log_2(Y_{ijk}) = \mu + \alpha_i + B_j + \varepsilon_{ijk}$, where i refers to different dye, j refers to independent sample preparation, k refers to individual array; $i = 1, 2, j = 1, 2$ and $k = 1, 2, 3$. The model was fit separately on the arrays from each condition. Y_{ijk} was the raw foreground probe intensity. α_1 and α_2 corresponded to the effect of different fluorescent dye. B_j, which is assumed to follow a normal distribution with mean 0 and variance $\sigma_B^2 (B_j \sim N(0, \sigma_B^2))$, represented the effect of independent *in vitro* transcription and labeling. $\varepsilon_{ijk} \sim N(0, \sigma^2)$ represented the residual error within technical replicates. We refer to the two variance components σ_B^2 and σ^2 as biological variance and technical variance for convenience, which estimate the variation introduced by sample preparation and array hybridization, respectively. The mixed effect model was fit using the *lmer* function from the lme4 library (Bates *et al*., 2008), and the restricted maximum likelihood estimators of the variance components were obtained.

2.6 Computation

All computation was performed using RedHat Enterprise Linux 4, 2.6 Kernel, R version 2.7.2 (R Development Core Team, 2008), and Bioconductor version 2.2 (Gentleman *et al*., 2004).

3 RESULTS

The Ag Spike dataset involves two groups of six arrays each from two independent sample preparations. The six arrays within each group contain three technical replicates and their corresponding dye-swapped pairs. Each array was hybridized with samples from two different conditions ('A' and 'B') (Fig. 1). The cRNA composition under each condition is identical to that in our previous Platinum Spike dataset except that the 26 cRNAs spiked with known absolute concentration in the Platinum Spike dataset were excluded here (Zhu *et al*., 2010). The Ag Spike dataset contains 1134 and 935 cRNAs up- and down-regulated, respectively, in condition A compared with B, with fold change values varying between 1.2× and 4×. The relative concentrations of another 3643 RNAs were kept unchanged between the two conditions. Similar to the Platinum Spike dataset, the Ag Spike dataset is balanced with respect to total labeled RNA amount and extent of up- and down-regulation for each experimental condition.

3.1 Detection of DEGs

To identify DEGs between the two compared conditions, 26 different routes were used to process the raw probe intensities in the Ag Spike dataset. These routes used different methods for background correction and normalization and used different test statistics for detecting DEGs. The performance of all routes was evaluated with ${}_rAUC_{0.05}$ ($0 \leq {}_rAUC_{0.05} \leq 1$, larger values = better performance), which measures the AUC value relative to the maximum (0.05) when the false positive rate (1 − specificity) is ≤ 0.05 (Supplementary Table S1).

Most routes give reasonable performance [median ${}_rAUC_{0.05} = 0.794$, median $TPR_{0.05}$ (true positive rate when

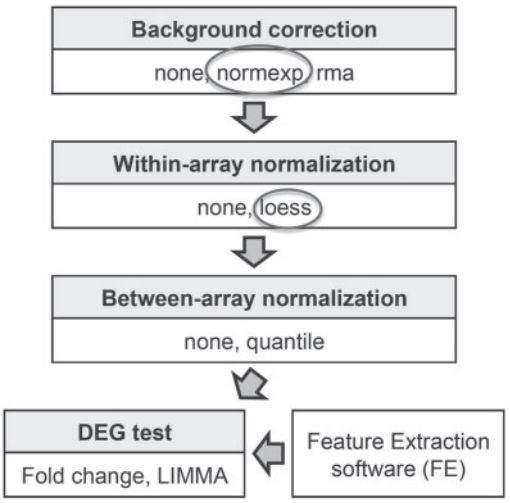

Fig. 2. Methods used at each step of analysis to identify DEGs. 'None' indicates that a given step was not performed. Methods superior to other methods are circled.

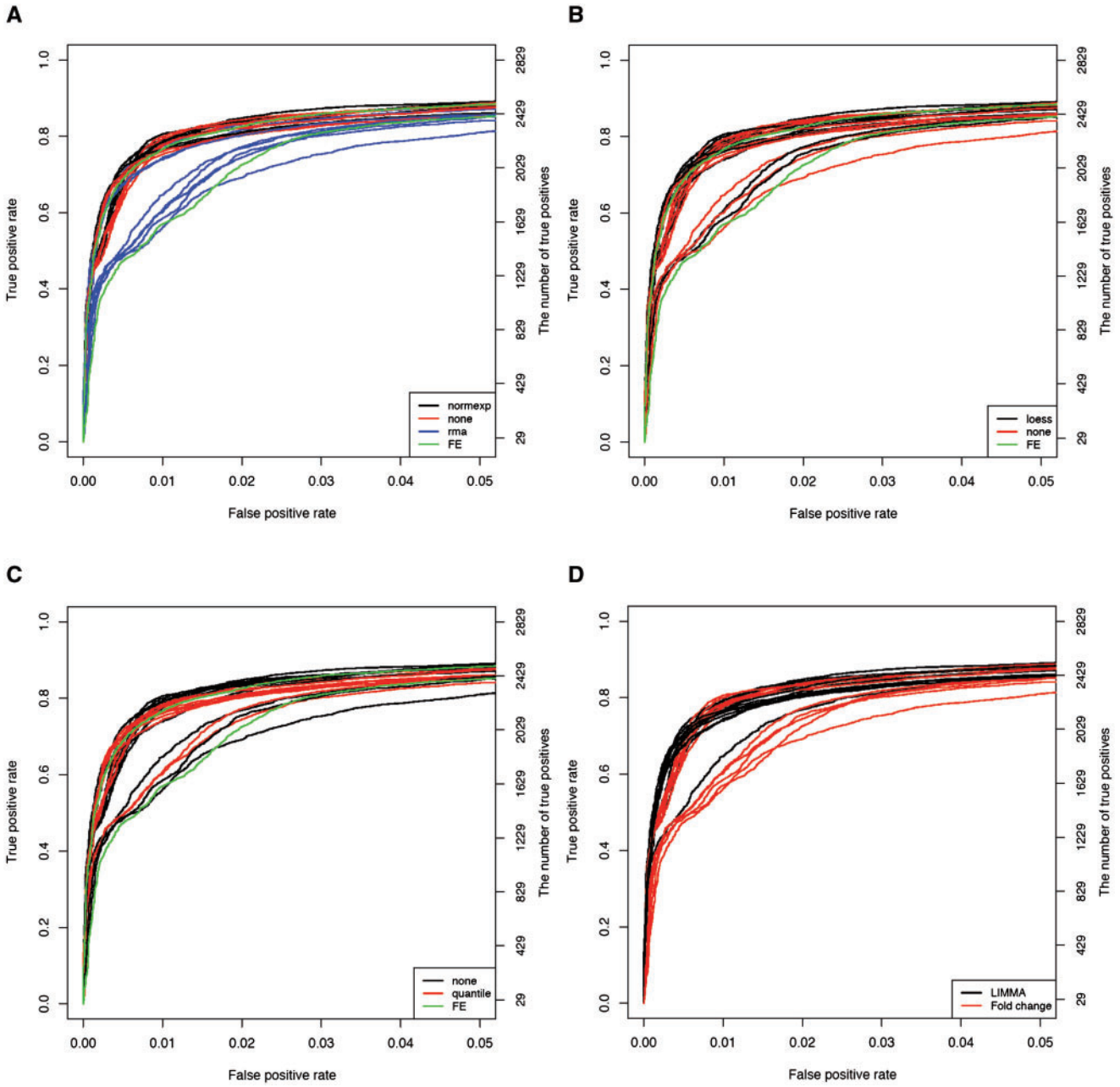

Fig. 3. Evaluation of DEG detection performance in the Ag Spike dataset. The ROC curves of the evaluated routes are highlighted in different colors according to the methods they used for background correction (**A**), within-array normalization (**B**), between-array normalization (**C**) and DEG testing (**D**).

false positive rate is $\leq 0.05) = 0.873$]. The best route for DEG detection in the Ag Spike dataset used *normexp* background correction, *loess* within-array normalization, and no between-array normalization and *LIMMA* for DEG testing ($_r\text{AUC}_{0.05} = 0.823$, $\text{TPR}_{0.05} = 0.890$). The performance of this route is comparable with the 97.8 percentile of routes evaluated in the Platinum Spike dataset based on $_r\text{AUC}_{0.05}$. In the background correction step, we found routes using *normexp* method perform consistently better than routes using other methods, while routes using the *rma* background method perform worse than most routes. Similarly, using *loess* within-array normalization leads to better performance than skipping within-array normalization (Figs 2 and 3). However, at the step of between-array normalization and DEG testing, no superior method is found as those corresponding to better average performance also result in large variation of performance (Fig. 3 and Supplementary Fig. S4).

3.2 Comparison with the Affymetrix dataset

Consistency across platforms has been a concern for microarray utilization (Li *et al.*, 2002; Tan *et al.*, 2003). In previous efforts to

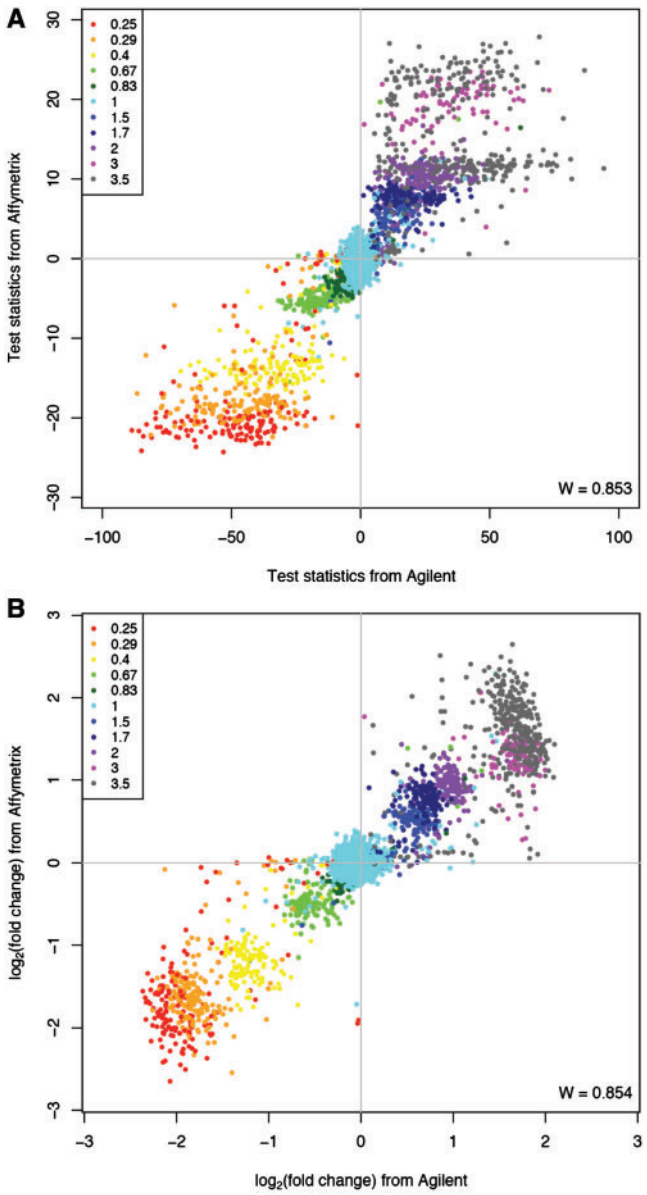

Fig. 4. The concordance between Agilent and Affymetrix platforms. The *x*- and *y*-axis correspond to the test statistics (**A**) or $\log_2$(fold change) (**B**) from the best route in Ag Spike dataset and Platinum Spike dataset, respectively. If one clone can be mapped to multiple probes/probesets, the median statistic or median $\log_2$(fold change) is plotted. Clones are highlighted in different colors based on their fold change levels (A versus B condition). The non-parametric Kendall's concordance coefficient (W) was calculated using the irr library in R.

address this issue, arrays from different platforms were hybridized with identical samples but with little information about the constituent RNAs (Irizarry *et al.*, 2005; Järvinen *et al.*, 2004; Patterson *et al.*, 2006; Tan *et al.*, 2003). The identical cRNA composition between the Ag Spike dataset and our previous Platinum Spike dataset allowed us to assess the consistency of the results from the two different platforms. We found that the overall performance of DEG detection is quite close between the best route from the Ag Spike dataset (*normexp* background correction, *loess* within-array normalization, no between-array normalization and *LIMMA* for DEG detection) and the best route from the Platinum Spike dataset (*gcrma-reb* background correction, *scaling* probe normalization among all arrays, using only PM probe intensities for *medianpolish* summarization, *vsn* probe set normalization among technical replicates and using *CyberT* for DEG detection) ($_rAUC_{0.05} = 0.823$ and 0.848, respectively). With false positive rate ≤ 0.05, 89% of the probes assigned to clones with distinct amounts in the two conditions are correctly detected as differentially expressed in the Ag Spike dataset, comparable with the top $TPR_{0.05}$ of 88% in the Platinum Spike dataset. We then focused on the 4935 clones that have assigned probes/probesets in both platforms and compared the concordance between test statistics from the two platforms using the best route from each. We found the test statistic values from the two platforms agree well with each other (Kendall's concordance coefficient $W = 0.853$, Fig. 4A). We also observed high concordance when comparing the $\log_2$(fold change) between conditions A and B of the two routes (Kendall's concordance coefficient $W = 0.854$, Fig. 4B). These results suggest that high concordance between the two platforms can be achieved when hybridization and data analysis are properly performed, consistent with results from the MAQC study (Patterson *et al.*, 2006).

4 DISCUSSION

We have generated the first wholly defined control dataset for the Agilent platform including over 2000 DEGs and 3600 non-DEGs. Using this control dataset, we compared a total of 26 analysis routes for detecting DEGs as a first-pass evaluation of the many existing analysis methods for dual-channel microarrays. The preprocessing steps of *normexp* background correction and *loess* within-array normalization performed better than other compared methods and were able to correct for technical errors that had resulted in overall higher intensities in a single array. We found that the results from the Ag Spike dataset were highly consistent with results from the Platinum Spike dataset, whose cRNA composition is identical to the current dataset but which was performed using the Affymetrix platform, when best routes from each analysis were compared. The Ag Spike dataset should prove a valuable community resource for further evaluation of the Agilent platform and for analysis of dual-channel microarrays in general.

ACKNOWLEDGEMENTS

This study utilized the high-performance computing cluster at the Center for Computational Research at SUNY Buffalo.

Funding: National Institutes of Health (grant R03 LM008941 to M.S.H.).

Conflict of Interest: none declared.

REFERENCES

Altschul,S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Bates,D. *et al.* (2008) lme4: linear mixed-effects models using S4 classes. R package version 0.999375-28. Available at http://cran.r-project.org/web/packages/lme4/.

Chen,Z. *et al.* (2007) A distribution free summarization method for Affymetrix GeneChip(R) arrays. *Bioinformatics*, **23**, 321–327.

Choe,S.E. *et al.* (2005) Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biol.*, **6**, R16.

Churchill,G.A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.*, **32**, 490–495.

Duggan,D.J. *et al.* (1999) Expression profiling using cDNA microarrays. *Nat. Genet.*, **21**, 10–14.

Gentleman,R. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.

Git,A. *et al.* (2010) Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA*, **16**, 991–1006.

Hochreiter,S. *et al.* (2006) A new summarization method for affymetrix probe level data. *Bioinformatics*, **22**, 943–949.

Irizarry,R.A. *et al.* (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.

Irizarry,R.A. *et al.* (2005) Multiple-laboratory comparison of microarray platforms. *Nat. Methods*, **2**, 345–350.

Järvinen,A.-K. *et al.* (2004) Are data from different gene expression microarray platforms comparable? *Genomics*, **83**, 1164–1168.

Kerr,K. (2007) Extended analysis of benchmark datasets for Agilent two-color microarrays. *BMC Bioinformatics*, **8**, 371.

Li,J. *et al.* (2002) Differential gene expression patterns revealed by oligonucleotide versus long cDNA Arrays. *Toxicol. Sci.*, **69**, 383–390.

Liu,S. *et al.* (2011) A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. *Nucleic Acids Res.*, **39**, 578–588.

Lockhart,D.J. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.

McCall,M.N. and Irizarry,R.A. (2008) Consolidated strategy for the analysis of microarray spike-in data. *Nucleic Acids Res.*, **36**, e180.

Patterson,T.A. *et al.* (2006) Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat. Biotechnol.*, **24**, 1140–1150.

Pearson,R. (2008) A comprehensive re-analysis of the Golden Spike data: towards a benchmark for differential expression methods. *BMC Bioinformatics*, **9**, 164.

R Development Core Team (2008) R: a language and environment for statistical computing. Available at http://www.r-project.org/.

Ritchie,M.E. *et al.* (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, **23**, 2700–2707.

Schena,M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.

Schuster,E. *et al.* (2007a) Correcting for sequence biases in present/absent calls. *Genome Biol.*, **8**, R125.

Schuster,E. *et al.* (2007b) Estimation and correction of non-specific binding in a large-scale spike-in experiment. *Genome Biol.*, **8**, R126.

Sing,T. *et al.* (2005) ROCR: visualizing classifier performance in R. *Bioinformatics*, **21**, 3940–3941.

Smyth,G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, 3.

Smyth,G.K. and Speed,T. (2003) Normalization of cDNA microarray data. *Methods*, **31**, 265–273.

Tan,P.K. *et al.* (2003) Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res.*, **31**, 5676–5684.

Turro,E. *et al.* (2007) BGX: a Bioconductor package for the Bayesian integrated analysis of Affymetrix GeneChips. *BMC Bioinformatics*, **8**, 439.

Wang,Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.

Willenbrock,H. *et al.* (2009) Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. *RNA*, 15, 2028–2034.

Zahurak,M. *et al.* (2007) Pre-processing agilent microarray data. *BMC Bioinformatics*, **8**, 142.

Zhu,Q. *et al.* (2010) Preferred analysis methods for Affymetrix GeneChips. II. An expanded, balanced, wholly-defined spike-in dataset. *BMC Bioinformatics*, **11**, 285.