both methods have a similar performance. In contrast, the cross-validation used in our paper was the 'leave-a percent-out', where a percent of the proteins with functional annotations is assumed unclassified. In this way, we conclude that the globally consistent method results in a noticeable performance increase over a local method based on a majority rule.

In summary, the authors propose a different functional assignment method and use a different cross-validation method than those used in our paper. We conclude, therefore, that further analysis of the performance of global versus local method is warranted before generalizable conclusions about the superiority of one approach over another can be made.

Reproducibility Probability Score—incorporating measurement variability across laboratories for gene selection

To the editor:

In the September issue, a paper entitled "The MicroArray Quality Control (MAQC) project shows interplatform reproducibility of gene expression measurements" (Shi, L. *et al.*, *Nat. Biotechnol.* **24**, 1151–1161, 2006) authored by us and others highlighted the need for a statistical metric to account for interlaboratory measurement variability in the selection of differentially expressed genes from microarray data. Here, we describe a novel metric (**Supplementary Methods 1** online) called the Reproducibility Probability Score (RPS), which is computed from gene expression data from a single laboratory. A gene with a higher RPS is evidence for differential expression that is more reproducible by other laboratories. We also provide a free, open source program (http://biocomp.bioen.uiuc.edu/rps) for computing the RPS to identify differentially expressed genes. Currently, the RPS program is capable of analyzing data from five commonly used microarray platforms—the Human Genome Survey Microarray v2.0 (Applied Biosystems, Foster City, CA, USA), the HG-U133 Plus 2.0 GeneChip (Affymetrix, Santa Clara, CA, USA), the Whole Human Genome Oligo Microarray G4112A (Agilent, Palo Alto, CA, USA), the CodeLink Human Whole Genome (GE Healthcare, Chalfont St. Giles, UK) and the Human-6 BeadChip 48K v1.0 (Illumina, San Diego)—and it can be extended to analyze other microarray platforms.

The RPS for a gene is defined as the probability that this gene is selected as being differentially expressed from the data generated by a typical laboratory. A typical laboratory is either the original laboratory that generated the microarray data or a hypothetical laboratory that prudently follows the same protocol to study the same biological materials as the original laboratory. To compute the RPS for a gene, the user needs to choose a traditional gene selection procedure, denoted by $\Theta > \theta$, where Θ is a statistic, such as the *P* value (or its inverse, if > in the gene selection procedure is interpreted literally), or a set of statistics, and θ is its corresponding threshold(s). The RPS for a gene is:

$$\mathrm{RPS} = \mathit{Prob}\ (\text{this gene is selected by a typical laboratory}) = E_k\{I\,[\Theta_k > \theta]\} \tag{1}$$

where k is the index of typical laboratories ($k = 0,1,2,\ldots$), with $k = 0$ denoting the original laboratory and $k > 0$ denoting the hypothetical laboratories. $E_k\{\bullet\}$ is the expectation over k. $I[\bullet]$ is the 0–1 indicator function, and Θ_k is the user-chosen metric computed from the data from the k^{th} laboratory. If there were a perfect correlation in the interlaboratory measurements, equation (1) would reduce into:

$$\mathrm{RPS} = \mathrm{I}\,[\Theta_0 > \theta] \tag{2}$$

which is identical to gene selection based on the user-specified procedure on the data from the original laboratory.

The RPS program uses simulation to generate data for the hypothetical laboratories in the computation of the RPS. Two sets of data are used in the simulation. The first set is the data generated from the original laboratory, also referred to as the new data. The second set is a reference data set. Currently, the RPS program uses the data from the Shi *et al.* as the reference data set, and it provides an option for the use of other reference data. The reference data and the new data have to be generated on the same microarray platform, and the reference data have to be generated by more than one laboratory. However, the reference data do not have to be generated by the same laboratory or originate from the same biological samples as the new data.

The workflow of the RPS program is as follows (**Fig. 1** and **Supplementary Methods 2** online). First, the RPS program applies a mixed-effects model to the reference data to estimate interlaboratory correlation for each probe (set). By default the MAQC data are preloaded as the reference data, but if the new data to be analyzed come from a microarray platform with no preloaded data, the user should provide appropriate reference data. Second, the RPS program reads in the new data to be analyzed. It uses the new data together with the estimated interlaboratory correlations to simulate the data for the hypothetical laboratories. Finally, the program uses the new data and the simulated data to compute an RPS for each gene, and ranks the genes by their RPS values.

To demonstrate the merit of the RPS algorithm, we have generated microarray data from five colorectal adenocarcinomas and matched normal colonic tissues. The RNA was first hybridized onto Affymetrix HG-U133-Plus-2.0 arrays in the Stanford Genome Technology Center (SGTC; Palo Alto, CA, USA). The RPS program and some other commonly used programs or procedures were used to identify differentially expressed genes. The other programs and procedures included MAANOVA[1], BayesAnova[2], FDR (Benjamini & Hocheberg procedure[3] on the two sample *t*-test), *P* value from the two sample *t*-test (computed by dChip[4]) and fold-change.

The same biological materials were then processed for hybridization in a different laboratory (the PAN facility on the Stanford

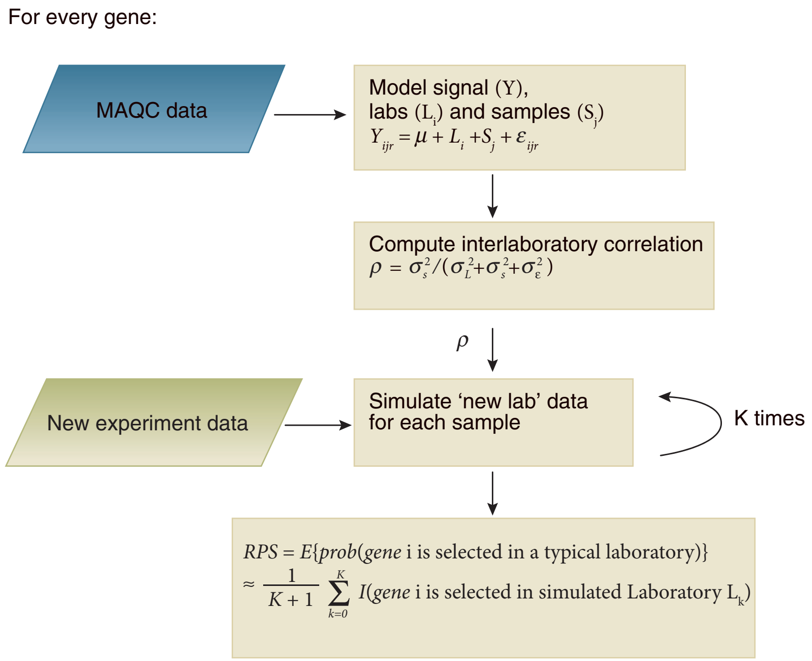
For every gene:
MAQC data
Model signal (Y), labs (L_i) and samples (S_j)
$Y_{ijr} = \mu + L_i + S_j + \varepsilon_{ijr}$
Compute interlaboratory correlation
$\rho = \sigma_s^2/(\sigma_L^2+\sigma_s^2+\sigma_\varepsilon^2)$
ρ
New experiment data
Simulate 'new lab' data for each sample
K times
$RPS = E\{prob(gene\ \text{i is selected in a typical laboratory})\}$
$\approx \frac{1}{K+1}\sum_{k=0}^{K} I(gene\ \text{i is selected in simulated Laboratory } L_k)$

Figure 1 Flow chart for the computation of the RPS for every probe (set). The 'MAQC data' and the 'New experiment data' represent data from completely different laboratories and different samples.

main campus) to independently check the reproducibility of the results obtained from the various gene selection programs applied on the SGTC data. The differentially expressed genes selected by the RPS program were much more reproducible than those selected by any of the other programs or procedures tested (**Supplementary Data 1** online). This result reflects the fact that none of the other programs or procedures considers the interlaboratory measurement variability in identifying differentially expressed genes. In another empirical investigation, we also confirmed that reproducibility of the genes selected by the RPS program was higher than those of other procedures (**Supplementary Data 2** online).

In summary, we provide the RPS program for selection of differentially expressed genes. It operates on a single laboratory's data and aims to deliver reproducible results (**Supplementary Data 3** online).

ACKNOWLEDGMENTS
The authors thank Douglas Simpson, Rex Gaskins, Walter Liggett and Alan Huebner for useful discussions. X.H. is supported by National Science Foundation DMS-0604229. H.J. is supported by National Institutes of Health (NIH) R21-CA109190, K08-CA96879, and by the Reddere Foundation. R.W.D. is supported by NIH P01 HG000205. S.Z. is supported by the Stowers Institute for Medical Research.

Note: Supplementary information is available on the Nature Biotechnology website.

***Guixian Lin*[1]*, Xuming He*[1]*, Hanlee Ji*[2]*, Leming Shi*[3]*, Ronald W Davis*[4] *& Sheng Zhong*[5]**

[1]Department of Statistics, University of Illinois at Urbana-Champaign, 725 S. Wright St., Champaign, Illinois 61820, USA. [2]Department of Medicine, Stanford University School of Medicine, 318 Campus Drive, Stanford, California 94305, USA. [3]National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Rd., Jefferson, Arkansas 72079, USA. [4]Departments of Biochemistry and Genetics, Stanford University School of Medicine, 318 Campus Drive, Stanford, California 94305, USA. [5]Departments of Bioengineering and Statistics and Computer Science and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1304 W. Springfield Ave., Urbana, Illinois 61801, USA.
e-mail: szhong@uiuc.edu

1. Kerr, M.K., Martin, M. & Churchill, G.A. *J. Comput. Biol.* **7**, 819–837 (2000).
2. Baldi, P. & Long, A.D. *Bioinformatics* **17**, 509–519 (2001).
3. Benjamini, Y. & Hocheberg, Y. J. *Royal Stat. Soc. B*, **57**, 289–300 (1995).
4. Li, C. & Wong, W.H. *Proc. Natl. Acad. Sci.* **98**, 31–36 (2001)