SUPPLEMENTARY INFORMATION ON MAQC STUDY

Contents

MAQC web site: http://edkb.fda.gov/MAQC/

A. Goals of the MicroArray Quality Control (MAQC) Project

- Establish a set of reference RNA samples for use within the MAQC project and more importantly for use by the scientific community;
- Generate a large collection of reference data sets using multiple microarray platforms across multiple laboratories per platform;
- Generate alternative measurements using other gene expression technologies for a large subset of genes;
- Measure relative accuracy as defined by titrated mixtures of the two reference RNA samples;
- Measure basic performance aspects of microarrays such as repeatability and reproducibility using a common set of genes across multiple platforms;
- Develop a framework for conducting cross-platform mapping based on probe sequence mapping to the RefSeq database and the AceView database;
- Develop quality control metrics and thresholds for objectively assessing the performance achievable by various microarray platforms;
- Compare the concordance of expression measurements to other microarray platforms and alternative technologies (*e.g*., TaqMan®, QuantiGene®, and StaRT-PCR™) based on a pre-defined list of "common genes" that share a common reference sequence to which the probes were designed;
- Evaluate the advantages and disadvantages of various data analysis methods with the intention of reaching consensus on microarray data analysis;
- Promote the use of reference RNA samples and reference data sets by the large scientific community for quality control and performance validation in order to identify and avoid potential procedural failures;
- Make recommendations on the appropriate uses of the microarray technology.

B. MAQC Participants and Available Data Sets

Table S1. Platform Providers and Test Sites

No.	Platform	Protocol	Test Site	Number of Microarrays	Organization*	Representative
1	Applied Biosystems	One-Color Microarray	ABI_1	20	Applied Biosystems	Yongming A. Sun
			ABI_2	19	US Environmental Protection Agency	David J. Dix
			ABI_3	19	Vanderbilt University	Shawn Levy
2	Affymetrix	One-Color Microarray	AFX_1	20	Affymetrix	Chunmei Liu
			AFX_2	20	FDA/CDER	Karol L. Thompson
			AFX_3	20	Asuragen	Mike Wilson
			AFX_4[b]	*20*	US Environmental Protection Agency	J. Christopher Corton
			AFX_5[b]	*20*	Novartis	Ron Peterson
			AFX_6[b]	*20*	UCLA/Cedars-Sinai	Charles Wang
3	Agilent	Two-Color Microarray	AGL_1[b]	*18*	Agilent	Patrick J. Collins
			AGL_2[b]	*18*	FDA/NCTR	Tucker A. Patterson
			AGL_3[b]	*20*	Cogenics/Clinical Data	Edward K. Lobenhofer
		One-Color Microarray	AG1_1	19	Agilent	Patrick J. Collins
			AG1_2	18	FDA/NCTR	Tucker A. Patterson
			AG1_3	19	Cogenics/Clinical Data	Edward K. Lobenhofer
4	GE Healthcare	One-Color Microarray	GEH_1	20	GE Healthcare	Richard Shippy
			GEH_2R[a]	*20*	UMass Boston	Michael J. Lombardi
			GEH_3	20	GenUs BioSystems	Scott R. Magnuson
5	Illumina	One-Color Microarray	ILM_1	19	Illumina	Shawn C. Baker
			ILM_2	20	UT Southwestern	Quan-Zhen Li
			ILM_3	20	Burnham Institute	Craig A. Hauser
6	NCI_Operon	Two-Color Microarray	NCI_1	20	NIH/NCI	Ernest S. Kawasaki
			NCI_2R[a]	13	FDA/NCTR	Tao Han
7	Eppendorf	One-Color Microarray	EPP_1	20	Eppendorf	Francoise de Longueville
			EPP_2	20	MD Anderson	Lajos Pusztai
			EPP_3	20	Cold Spring Harbor Laboratory	Eli Hatchwell
8	Applied Biosystems	TaqMan® Assays	TAQ	N/A	Applied Biosystems	Kathleen Y. Lee
9	Panomics	QuantiGene® Assays	QGN	N/A	Panomics	Yuling Luo
10	Gene Express	StaRT-PCR™ Assays	GEX	N/A	Gene Express/Ohio Medical University	James C. Willey
			TOTAL	***502***		

[a]Original MAQC data set replaced with a repeat data set in the main study.

[b]Data sets not included in this publication. The two-color AGL data is presented in [Patterson, T.A. *et al*., *Nat. Biotechnol*. **24**(9), 2006]. The total number of microarrays includes 386 hybridizations that were included in this article and 116 hybridizations that were not included.

Table S2. Data Analysis Sites

No.	Organization	Representative
1	Biogen Idec	Lisa J. Croner
2	Expression Analysis	Wendell Jones
3	FDA/NCTR	Leming Shi
4	Harvard Univ./Children's Hospital	Zoltan Szallasi
5	NIH/NCBI	Damir Herman
6	National Institute of Standards and Technology	Walter Liggett
7	SAS	Russ Wolfinger
8	Stanford Univ.	Hanlee Ji
9	UIUC	Sheng Zhong
10	Univ. of Massachusetts-Boston	Roderick V. Jensen
11	ViaLogy	Cecilie Boysen

Table S3. Data Replaced or Removed from Main Study Analysis

Data Name	Reason for Removal or Replacement	Number of Microarrays
ABI_2_A5	RT control probe LYS_1 showed 20X higher signals compared with mean of other sample A indicating unsuccessful RT reaction	1
ABI_3_D2	Refrigerator failure caused microarray to be warm and dried 'swirled fluid' occurred on the microarray	1
AG1_1_A1	Determined as outlier using metrics in Agilent's Feature Extraction QC Report	1
AG1_2_A3	Determined as outlier using metrics in Agilent's Feature Extraction QC Report	1
AG1_2_D2	Determined as outlier using metrics in Agilent's Feature Extraction QC Report	1
AG1_3_B3	Determined as outlier using metrics in Agilent's Feature Extraction QC Report	1
AGL_1_B5	Determined as outlier using metrics in Agilent's Feature Extraction QC Report	1
AGL_1_D1	Determined as outlier using metrics in Agilent's Feature Extraction QC Report	1
AGL_2_A1	Determined as outlier using metrics in Agilent's Feature Extraction QC Report	1
AGL_2_C4	Determined as outlier using metrics in Agilent's Feature Extraction QC Report	1
ILM_1_C3	Low IVT yield and poor Bioanalyzer trace; Target was not hybridized	*0*
ILM_2_D3	Low IVT yield; Repeat sample named ILM_2_D6; counted n Table S1.	*1*
ILM_2_D5	Low IVT yield; Repeat sample named ILM_2_D7; counted n Table S1.	*1*
NCI_2R_A3	Indicated by the site as an outlier microarray.	1
GEH_2 Set	Data generated in August 2005 used different protocol than other GEH test sites. Replaced by GEH_2R data generated in February 2006 at same site using corrected protocol.	20
NCI_2 Set	Data generated in October 2005 had problems with the hybridization buffer. Replaced by NCI_2R data generated in November 2005 which includes 14 hybridizations: sample A = 4 hybridizations, sample B = 4 hybs, sample C = 3 hybs and sample D = 3 hybs.	20
NCI_3 Set	Data generated in September-October 2005 had problems with the hybridization buffer. Replaced by NCI_3R data generated December 2005-January 2006 which includes 10 hybridizations: 4 pairs of dye-swap hybs with A and B, plus 2 self-self hybs with sample A.	20
	TOTAL	***71***[a]

[a]The total number of microarrays includes four AGL hybridizations that were not analyzed in this paper, 11 hybridizations that were replaced for quality issues and 60 hybridizations that were removed or replaced for protocol issues.

Table S4. Other Available Data Sets

Platform	Code	Protocol	Test Sites	Number of Samples	Number of Replicates	Number of Microarrays Per Test Site
NCI_ Operon	NCI	Two-Color Microarray	NCI_4: Harvard Univ.	2	5	10
			NCI_3R: FDA/CBER	2	5	10
CapitalBio_ Operon	BIO	Two-Color Microarray	BIO_1: CapitalBio	2	10	20
	BIO1	One-Color Microarray	BIO1_1: CapitalBio	2	5	10
Operon	OPN	Two-Color Microarray	OPN_1: Operon	2	5	10
NMC_ Operon	NMC	Two-Color Microarray	NMC_1: Norwegian Microarray Consortium	2	5	10
TeleChem	H25K	Two-Color Microarray	H25K_2: Yale Univ.	2	15	30
	H25K1	One-Color Microarray	H25K1_1: TeleChem	4	5	10
			H25K1_2: Yale Univ.	4	5	10
			H25K1_3: Wake Forest Univ.	4	5	10
					TOTAL	***130***

Table S4-A. "*Tumor*" Data Included in the MAQC Study[a]

(One microarray platform at two laboratories in Stanford University)

Manufacturer	Code	Protocol	Platform	Number of Probes	Number of Test Sites	Number of Samples	Number of Replicates	Total Number of Microarrays
Affymetrix	AFX	One-Color Microarray	HG-U133 Plus 2.0 GeneChip®	54,675	2	2	5	**20**

[a]Tumor_Stanford_Lab1 and Tumor_Stanford_Lab2. T = Tumor (colon adenocarcinoma), N = Normal (normal colon tissue, patient matched). The tumor data set was analyzed in Lin, G., He, X., Ji H., Shi, L., Davis, R.W. and Zhong, S. *Nature Biotechnology*, **24**(10), 2006.

Table S4-B. "Rat *Toxicogenomics*" Validation Data Included in the MAQC Study

Manufacturer	Code	Protocol	Platform	Number of Probes	Number of Test Sites	Number of Samples	Number of Replicates	Total Number of Microarrays
Applied Biosystems	ABI	One-Color Microarray	Rat Genome Survey Microarray	26,857	1	6	6	36
Affymetrix	AFX	One-Color Microarray	Rat Genome 230 2.0 GeneChip®	31,099	2	6	6	72
Agilent	AG1	One-Color Microarray	Whole Rat Genome Oligo Microarray, G4131A	43,628 (41,071 – GeneSpring[a])	1	6	6	36
GE Healthcare	GEH	One-Color Microarray	Rat Whole Genome Bioarray, 300031	35,129	1	6	6	36
							TOTAL (Toxicogenomics)	**180**

[a]Data from replicating spots were averaged within GeneSpring software to generate a single value for each unique probe. The rat toxicogenomics data set was analyzed in Guo, L. *et al*. *Nat. Biotechnol.* **24**(9), 1162-1169(2006).

The total number of microarrays used in the MAQC project: 1,329

1. Official:	573 (502 from Table S1 and 71 from Table S3)
2. Additional:	130 (Table S4)
3. Tumor:	20 (Table S4-A)
4. Rat Toxicogenomics:	180 (Table S4-B)
5. Pilots (I* and II**):	426

*__MAQC Pilot-I__ (RNA Sample Selection): **160** microarrays (four human platforms).

AFX (2 sites):	40
AGL (2 sites):	60
GEH (2 sites):	40
ILM (1 site):	20

MAQC Pilot-I was designed to select two RNA samples for the MAQC main study. Four distinct RNA samples from three commercial vendors were tested on four microarray platforms (AFX, AGL, GEH, and ILM). Except for ILM, each platform was tested at two test sites. For each one-color platform, each RNA sample was tested in five replicates at each test site. Two RNA samples, Stratagene's Universal Human Reference RNA (UHRR) and Ambion's Human Brain Reference RNA (HBRR), designated as MAQC sample A and sample B, respectively, were selected during the second MAQC face-to-face meeting held at FDA/CDER, Rockville, Maryland on May 2-3, 2005 based on criteria including

1. Availability in large quantity from a single batch;
2. Accessibility from commercial sources;
3. High quality;
4. Wide gene presence;
5. Large fold changes for a number of genes for the sample pair;
6. Reproducibility in production.

Consensus on RNA sample selection was reached after Pilot-I data were analyzed, presented, and debated.

__MAQC Pilot-II__ (RNA Sample Titration): **266 microarrays (six human platforms).

AFX (1 site):	45
AGL (1 site)	45
AG1 (1 site):	45
GEH (1 site):	46
ILM (1 site):	51
Anonymous (1 site):	34

MAQC Pilot-II was designed to determine the appropriate mixing ratios of sample A and sample B in order to create the titration pools. A decision to include titration mixtures in the MAQC main study was made during the MAQC first face-to-face meeting at FDA/NCTR, Jefferson, Arkansas on February 11, 2005. The experimental design of Pilot-II was listed below. After reviewing the data, the MAQC consortium decided on the 75/25 and 25/75 (A/B) ratios for creating the MAQC samples C and D, respectively.

No.	Sample A (%)	Sample B (%)	Number of Replicates[a]
1	100	0	6
2	99.5	0.5	3
3	99	1	3
4	95	5	3
5	90	10	3
6 (**C**)	**75**	**25**	3
7	50	50	3
8 (**D**)	**25**	**75**	3
9	10	90	3
10	5	95	3
11	1	99	3
12	0.5	99.5	3
13	0	100	6
Total number of arrays per site (manufacturer)			**45**

[a]The actual number of replicates may have varied from platform to platform. TaqMan (TAQ) and QuantiGene (QGN) were also used in Pilot-II study.

Note: Pilot-I and Pilot-II data will not be deposited in public databases until further notice.

C. Quality Assessments

The median length and total yields of amplified RNA (cRNA) were measured for each MAQC sample. These data are available in **Supplementary Table 1** online: MAQC cRNA Sizes and Yields. The purified cRNA size was analyzed using an RNA LabChip assay on the Agilent 2100 Bioanalyzer. The median size was determined by selecting the point (nt) which evenly divided (50%) the area in the electropherogram. Yields of the purified cRNA were determined by spectrophotometry (A260). Yields were reported as either the total amount of cRNA generated or as the normalized output per input of total RNA [cRNA(ug) / Total RNA(ng)] For example, a reaction generating 111.5 ug cRNA from 1000 ng input total RNA has a normalized output of 0.11 ug/ng.

D. Platform-Specific Protocols and Data Analysis Methods

Applied Biosystems Microarrays

Applied Biosystems Human Genome Survey Microarray is part of the Applied Biosystems Expression Array System (http://docs.appliedbiosystems.com/pebiodocs/04338853.pdf), which includes a 1700 Chemiluminescent Microarray Analyzer, labeling and hybridization chemistries, and supporting software. Digoxigenin-UTP labeled cRNA was generated from 1 μg of total RNA for each MAQC sample (A, B, C, D) using Applied Biosystems NanoAmp™ RT-IVT Labeling Kit (P/N 4365715) according to the manufacturer's protocol. Array hybridization, array processing, chemiluminescence detection, image acquisition, and analysis were performed using Applied Biosystems Chemiluminescence Detection Kit and Applied Biosystems 1700 Chemiluminescent Microarray Analyzer following manufacturer's protocol. These protocols are detailed in the Chemiluminescent Microarray Analyzer Chemistry Guide (P/N 4338853, http://docs.appliedbiosystems.com/pebiodocs/04338853.pdf), Chemiluminescent Detection Kit (P/N 4339627, http://docs.appliedbiosystems.com/pebiodocs/04339627.pdf), and Chemiluminescent Microarray Analyzer User Guide (P/N 4338852B).

The Expression Array System Software suite performs the auto-gridding, feature extraction, fluorescence normalization, and signal data generation. Each probe contains a universal 24-mer control probe (ICP) co-localized. The fluorescent signal of the ICP is used in the quantification process to normalize probe chemiluminescent signal within an array. The quantification data contain many distinct measurements for each probe, including the three basic measurements: Signal, S/N, and Flag values. The Signal value is the fully corrected, background subtracted measurement of chemiluminescent signal for gene expression values. The S/N value represents the ratio of signal above noise, or the measurement uncertainty of the probe signal, and can be used as a confidence level for the probe measurement. An S/N of 3 represents approximately a 99.95% confidence that the probe is detected above the background noise. In situations where the probe showed S/N < 1, the signal measurement is replaced with a 1 SDEV upper limit based on its probe signal SDEV. The Flag value is generated for each probe from a numeric code from a series of feature quality metrics that give a descriptive quality value. Probe signals for FLAG values equal or exceeding 8192 ($>= 2^{13}$) are considered missing values and imputed in data analysis. In addition, there are varieties of control probes included on the array for quality monitoring of various microarray processes. The between array quantile normalization was performed independently for each test site after removing the control probes. Detailed image analysis algorithms and data processing can be found in document http://docs.appliedbiosystems.com/pebiodocs/04367370.pdf. The probe sequences are available at http://www.pantherdb.org.

Affymetrix GeneChip® Microarrays

MAQC samples were processed following Affymetrix one-cycle sample preparation protocol (Please refer to GeneChip® Expression Analysis Technical Manual for details) and hybridized to Human Genome U133 Plus 2.0 array (catalog #900470, #900466 or #900467). Affymetrix GeneChip® Operating Software (GCOS) is used to automate the control of GeneChip® Fluidics Stations and Scanners. In addition, GCOS acquires data, manages sample and experimental information, and performs gene expression data analysis. The Probe Logarithmic Intensity ERror (PLIER) algorithm is used to produce a summary value for a probe set by accounting for

experimentally observed patterns for feature behavior and handling error appropriately at low and high abundance. Probe set signal intensity is transformed by adding a constant of 16 and then taking log2 transformation. We use routine QC parameters to monitor the quality of the experiment, which includes visual array inspection, background, scaling factor, noise, 3'/5' GAPDH and Actin ratios, and % Present calls. Probe sequence information is publicly available at http://www.affymetrix.com/support/technical/byproduct.affx?product=hg-u133-plus.

Agilent Microarrays

Data were generated for the MAQC project using both Agilent's One-Color Gene Expression Platform and Agilent's Two-Color Gene Expression Platform. For the one-color data sets (AG1) all four MAQC samples (A – D) were assayed, while for the two-color data sets (AGL) only samples A and B were assayed. Sample labeling, array processing and scanning, and data extraction were performed as described in either One-Color Microarray-Based Gene Expression Analysis – Protocol (AG1) or Two-Color Microarray-Based Gene Expression Analysis – Protocol (AGL). These protocols are available for download at http://www.chem.agilent.com/scripts/generic.asp?lpage=11617&indcol=N&prodcol=Y. The microarrays used were Agilent's Whole Human Genome Oligo Microarrays (P/N G4112A). Probe sequence information is publicly available at http://www.chem.agilent.com/cag/bsp/bsp_register.asp.

Labeled cRNAs were generated from 500 ng of total RNA for each of AG1 and AGL using Agilent's Low RNA Input Linear Amplification Kit (P/N 5188-5339). The two-color protocol was modified such that 1.5 μg of labeled cRNA was hybridized per dye channel per microarray. Agilent's Stabilization and Drying Solution was used in processing all microarrays. Microarrays were scanned using Agilent's DNA Microarray Scanner BA (P/N G2565BA) and data extracted using Agilent's Feature Extraction software, version 8.5 (P/N G2567AA). AG1 data were processed using Agilent's GeneSpring® GX software, version 7.3 (P/N G1745). Data were transformed by setting all measurements less than 5.0 to 5.0. Data points that did not have detectable signal and those that represent microarray controls were labeled as Absent, those representing either non-uniform or saturated features were labeled as Marginal, and all remaining data points were labeled as Present. All data points were median scaled using the median signal intensity value for data points labeled as Present. For AGL data points that did not have detectable signal in both dye channels and those that represent microarray controls were labeled as Absent, those representing either non-uniform or saturated features in either channel were labeled as Marginal, and all remaining data points were labeled as Present.

Eppendorf DualChip® Microarrays

Each DualChip® consists of two identical microarrays on the same slide, with two separate hybridization frames. This DualChip®-concept enables two complete gene expression experiments to be run on the same slide, using a one-color labeling technique. The DualChip® array used for the MAQC consortium contains capture probes allowing the expression analysis for 294 human genes spotted in triplicate. A reverse transcription with indirect labeling was performed using 10 μg of total RNA. The internal standard mixture was used for quantification/ normalization and estimation of experimental variation [de Longueville F, *et al. Biochem. Pharmacol.* **64**, 137-49, 2002; Chen & Bittner. *J. Biomedical Optics* **2**, 364-374, 1997]. The DualChip® hybridization was carried out overnight (16 hours) at 60°C according to the DualChip® instruction manual. The detection was

performed by using a Cy3-conjugated anti-biotin IgG (Jackson Immuno Research Laboratories, West Grove, PA) on biotinylated cDNA.

Using the DualChip® evaluation software, the fluorescence intensities of each DNA spot (average intensity of each pixel present within the spot) was calculated using local mean background subtraction. Very bright element intensities (saturated signals, highly expressed genes) and undetected elements were deemed unsuitable for accurate quantitative analysis. The reverse transcription and hybridization efficiency were controlled by the internal standards spiked previously in the samples. First, the values were corrected using a factor calculated from the intensity ratios of the internal standards in the reference and in the experimental samples. A second normalization was performed involving the average ratio from a set of housekeeping genes for which the expression level was effectively constant.

Within the microarray platforms participating to the MAQC project, Eppendorf platform is the only one providing low density arrays. This characteristic induces by itself major differences in the resulting data structure. Analysis procedures (normalization, …) for high density arrays rely on assumptions that are not valid for low density arrays. Eppendorf has developed specialized data analysis scheme treating the scanning, the background correction and normalization in a suited framework for low density arrays generating array to array gene expression ratio. The data sets submitted are divided in two groups. The first data set provides gene expression ratio comparing two samples and constitutes the standard result for the Eppendorf platform. In addition, normalized intensities intra-sample have been submitted to allow the analysis of reproducibility of intensities of detected (but not saturated) genes intra-sample. Due to the difference of scanner technologies used in the different sites and to our standardized scanning procedure, the detection concordance analysis can not be evaluated from this data set. The small gene subset, the intensive use of detection and saturation flags among the 3 sets of data produced for each array through the standardized triple scanning procedure, the background correction producing negative values and scanning procedure itself render invalid the direct comparison of Eppendorf data with the other platforms. To obtain relevant comparison including all the microarray platforms, the subset of genes has been reduced to the 294 spotted on the Eppendorf DualChip®. For each individual computation, this subset of genes has been limited to the "Present" flagged genes. Furthermore, reproducibility of ratio was computed on ratio compliant with the DualChip® data analysis (B1/A1, B2/A2, B3/A3,…) for all microarray platform.

GE Healthcare CodeLink™ Microarrays

MAQC total RNA samples were processed following GE Healthcare's CodeLink™ iExpress Expression Assay Reagent Kit protocol (Please refer to CodeLink™ iExpress Expression Assay Regent Kit Technical Manual for details at www.codelinkbioarrays.com) and hybridized to CodeLink™ Human Whole Genome bioarrays (catalog #300026). All arrays were scanned on the GenePix 4000B scanner using manufacturer recommended settings of 600 PMT, 100% laser power, and 5 micron resolution. The resulting images were processed through the CodeLink Expression Analysis v4.1 software which automatically grids and quantifies each feature while assigning spot quality metrics. These quality metrics are applied to each spot and are displayed both numerically and as a qualitative assignment. In the MAQC study, for all microarray platforms, genes which were not detected (*i.e.,* 'absent') were excluded from subsequent analyses. The spots which are below noise on the GE Healthcare platform are flagged by the software package as "L" and

represent cases where the spot mean intensity is less than local background median plus 1.5 standard deviations of local background. Details on the CodeLink™ detection algorithms can be found at the www.codelinkbioarrays.com (http://www1.amershambiosciences.com/APTRIX/upp00919.nsf/Content/WD%3AImproved+method%28275651129-B500%29?OpenDocument&hometitle=WebDocs).

The CodeLink™ image quantification algorithms have been empirically optimized using concentration response curves [illustrated in Figure 2 of Tong, W. *et al*., *Nat. Biotech.* **24**(9), 2006]. The lower limit of detection for the CodeLink™ platform is at least 50 fM, with 1:2,048,000 mass ratio 'spiking' into cRNA, with 99.2% of negative controls flagged below the detection limit. In other words, the detection algorithms for CodeLink™ has been tested down to or is approximately been optimized such that the level where signal is indistinguishable from noise which correlates to the non-specific negative control values and concentration response curves. It is recommended that these detection spot quality metrics are utilized since values below noise are unreliable and reduce correlations between platforms [Shippy, R., *et al., BMC Genomics* **5**, 61, 2004]. The CodeLink™ Expression Analysis Software performs the auto-gridding, feature extraction, fluorescence normalization, signal data generation, and assigns spot quality metrics. Probe sequence information, bioarray content, and custom array formatting options can be found within the CodeLink™ iCenter at www.gehealthcare.com/codelinkicenter.

Illumina Micorarrays

For each technical replicate, 200 ng of total RNA provided by the MAQC were amplified and labeled using the Illumina® TotalPrep™ RNA Amplification kit (Ambion; catalog #IL1791). Illumina Human-6 whole genome microarrays (Illumina; catalog # BD-25-101) were hybridized to 1.5 micrograms of labeled, amplified material, and then washed, stained, and scanned according to the protocol described in the Illumina Whole Genome Gene Expression for BeadStation Manual, Revision D. Scanning software was BeadScan 2.3.0.10. Data was processed using Illumina BeadStudio version 1.5.0.34 software using background subtraction and cubic spline normalization. Normalized hybridization intensity values were adjusted by adding a constant such that the lowest intensity value for any sample equaled 16. Individual sequences are available at http://www.illumina.com/General/Products/ArraysReagents/zip_files/Human_WG-6_sequence.zip

NCI_Operon Microarrays

Labeling of probes, hybridization of slides and scanning was done essentially as described [Petersen, D. *et al. BMC Bioinformatics* **6**, 63. 2005]. Slides were scanned using an Axon 4000B instrument at 10 micron resolution. All scans were performed at 100% laser power, but the PMT setting varied from slide-to-slide depending on the labeling and hybridization outcomes. Thus, the intensity of the signals among the slides will vary, but the Cy3 and Cy5 ratios will be much less variable. Images were processed and quantified using Genepix 4.0 software (Axon Instruments, Union City, CA). The data were filtered (or flagged) on the basis of signal levels and spot quality. The raw data were submitted to MAQC and further normalized by using LOESS. Probe sequences are available at http://omad.operon.com/download/index.php.

Applied Biosystems TaqMan® Assays

MAQC total RNA Samples A, B, C and D were reverse transcribed using random hexamers and Applied Biosystems cDNA Archive Kit to generate cDNA. There was no pretreatment of RNA

before cDNA preparation. Each TaqMan® Gene Expression Assay consists of two target specific PCR primers and a TaqMan FAM™ dye-labeled MGB hybridization probe in a pre-formulated mixture that contains the three oligonucleotides. The cDNA sample and assay reagents are combined together in a single reaction well and no further handling is required. Each TaqMan® Assay was run in four replicates for each RNA sample using 10ng total cDNA (measured as total input RNA) in a 10ul final volume. Assays were run with 2X Universal Master Mix (without UNG) on Applied Biosystems 7900 Fast Real-Time PCR System using universal cycling conditions (10 min. at 95^o C; 15 sec. at 95^o C; 1 min. at 60^o C for 40 cycles). POLAR2A was chosen as the reference gene and each replicate C_T was subtracted from the average POLAR2A C_T to give the log2 difference (ΔC_T), the normalized value. The ΔC_T of each replicate for each TaqMan® Assay was presented in the final data set as the normalized data. A raw C_T value of 35 (estimated as 5 copies) was set as the limit of detection in this study. Individual replicates with a C_T value >35 were considered not detected, or absent (A); replicates with $C_T < 35$ were considered detectable and identified as expressed or present (P). For ΔC_T calculations, we used C_T of 35 for any replicate that had $C_T > 35$. Fold changes between samples were calculated using the $\Delta\Delta C_T$ method. Detailed information on TaqMan® Gene Expression Assays Protocol can be found at http://docs.appliedbiosystems.com/pebiodocs/04333458.pdf. Individual TaqMan® Gene Expression Assay information can be found at: https://products.appliedbiosystems.com/ab/en/US/adirect/ab?JSESSIONID=Gtw2GJ2vxqyptZnQgXnvG8h41z3p4FyNb1Dy9pw5cFTNc39g1zhZ!-1961426624&cmd=ABGEKeywordSearch&catID=601267

Panomics QuantiGene® Assays

The QuantiGene® assays were performed according to the procedure of QuantiGene® Reagent System (Panomics), which was previously described in detail. Briefly, 10 μl of starting total RNA (500 ng) from sample A, B, C or D was mixed with 40 μl of Lysis Mixture (Panomics), 40 μl of Capture Buffer (Panomics), and 10 μl of target gene-specific probe set (CE, 1.65 fmol/μl; LE, 6.6 fmol/μl; BL, 3.3 fmol/μl). Each sample mixture was then dispensed into an individual well of a Capture Plate (Panomics). The Capture Plate was sealed with foil tape and incubated at 53°C for 16-20 h. The hybridization mixture was removed and the wells were washed 3 times with 250 μl of Wash Buffer (0.1X SSC, 0.03% Lithium Lauryl Sulfate). Residual Wash Buffer was removed by centrifuging the inverted Capture Plate at 1000x g. Signals for the bound target mRNA were developed by sequential hybridization with branched DNA (bDNA) amplifier, and alkaline phosphatase-conjugated label probe, at 46°C for 1 hour each. Two washes with Wash Buffer were used to remove unbound material after each hybridization step. Substrate dioxetane was added to the wells and incubated at 46°C for 30 min. Luminescence from each well was measured using a Lmax microtiter plate luminometer (Molecular Devices). Three replicate assays measuring RNA directly (independent sampling $n = 3$) were performed for all described experiments. Genomic DNA contamination in the RNA sample, if there is any, does not affect the QuantiGene® assay, since it remains doubled-stranded throughout the entire procedure and thus can not hybridize to the probe sets at the temperature used in the assay.

The QuantiGene® assays of 244 genes were performed for MAQC samples A, B, C, D. For all samples, background signals were determined in the absence of RNA samples and subtracted from signals obtained in the presence of RNA samples. Because QuantiGene® assay measures RNA directly, no data normalization against a reference gene is required in the data analysis. The

presence and absence call is determined by limit of detection (LOD) of the assay, where LOD = Background + 3SD of Background. If at least two samples out of A, B, C, D has signal below LOD in a gene, we call the gene absent. To determine gene expression fold change in Sample A versus Sample B, the fold change (FC) was calculated using formula log2 FC = $\log2(S_A/S_B)$, where S_A represents the assay signal for a target gene in sample A and S_B represents the assay signal for the target gene in sample B. A gene is considered for fold change analysis if the signal in both sample A and sample B passes the LOD.

Gene Express StaRT-PCR™ Assays

Standardized RT (StaRT)-PCR™ is a quantitative competitive template PCR platform that provides numerical quantification of transcript abundance values at endpoint of PCR over the full range observed in human tissues (more than six orders of magnitude). Following reverse transcription, cDNA molecules representing each gene are measured in relation to a known number of its respective internal standard cDNA molecules. In this MAQC project StaRT-PCR™ contributed to calibration of the UHRR and HBRR samples by generating numerical transcript abundance data that are directly comparable on a numerical basis to data collected by other investigators in the future. Quantification of cDNA molecules down to as few as 10 molecules/10^6 ACTB molecules through StaRT-PCR™ enabled determination of the relationship between number of molecules loaded and measurement variance due to stochastic sampling error. Inclusion of an internal standard in each measurement ensures 100% signal-to-target response in each measurement and provides a useful reference point to assess compression or expansion of signal-to-target response by other platforms. StaRT-PCR™ data are directly compared to QuantiGene® and TaqMan® data in another manuscript [Canales, R.D. *et al*., *Nat. Biotechnol.* **24**(9), 2006].

E. Probe Mapping

Table S5. Summary of Probe Mapping per Gene Expression Platform

Platform	Length of mapped probe sequence[a] (nt)	Number of probe sequences analyzed[b]	Number of probes that met mapping criteria[c] (percent of all probes, %)		Number of RefSeq NM Accessions mapped to probes[e]	Number of Entrez genes ID's mapped to probes via		Number of genes, not in Entrez, mapped to probes via AceView[d,f,g]
			RefSeq	AceView[d]		RefSeq[e]	AceView[d,f]	
ABI	60	32,878	18,547 (56.4)	25,566 (77.8)	21,963	16,763	18,676	3,267
AFX	25	54,675	24,694 (45.2)	44,693 (81.7)	21,318	15,965	18,911	10,129
AG1	60	41,000	22,677 (55.3)	32,024 (78.1)	21,890	16,493	18,051	4,055
GEH	30	53,423	16,881 (31.6)	43,540 (81.5)	20,230	15,429	16,984	18,408
ILM	50	47,282	20,140 (42.6)	31,229 (66.0)	22,161	16,990	18,797	8,666
NCI	39-70	35,235	21,555 (61.2)	29,396 (83.4)	20,987	15,899	17,641	1,411
EPP	161-513	294	285 (98.6)	285 (98.6)	315	285	290	0
QGN	183-2,671	245	234 (95.5)	234 (95.5)	253	233	237	0
GEX[h]	N/A	205	N/A	N/A	203	203	203	N/A
TAQ[h]	N/A	1,004	N/A	N/A	997	997	997	N/A
Union of six platforms[i]		264,493	125,216 (47.3)	206,448 (78.1)	23,971	18,114	21,662	32,025
Intersection of six platforms[i]					15,615	**12,091**	13,327	9
Intersection of six platforms[i] *and TAQ*[h]						**906**		

[a]For the AFX platform, the length of each individual probe is given. For the QGN platform, the length of the intended target is given.

[b]The number of probes for which mapping was attempted may slightly differ from the number of probes arrayed (**Table 1**) because of the removal of control probes and replicate spots. For the AFX platform, the number of probe sets is given.

[c]Probes were mapped as described in the Methods section. An exact sequence match was required and probes that match more than one gene were excluded. For the AFX platform, there are generally 11 probes per probe set, and each probe was mapped individually. An exact match of 80% of the probes in a probe set was required for the probe set to qualify as a perfect match. All the mapping data supporting this table are available from supplementary materials online and the MAQC web site (http://edkb.fda.gov/MAQC/).

[d]AceView is a transcriptome database that combines RefSeq, GenBank and dbEST entries [Thierry-Mieg, D & Thierry-Mieg, J, Genome Biology **7** (Suppl 1):S12**,** 2006]. For the details on the AceView mapping, please refer to the supplementary materials online at ftp://ftp.ncbi.nlm.nih.gov/repository/acedb/MAQC/MaqcMapping2AceViewTranscripts.zip.

[e]The numbers in these columns illustrate the source of the common set of 12,091 genes represented on the six high-density microarray platforms which have an overlap of 906 genes with the TAQ platform. The data do not fully reflect the coverage of each platform because the degree to which RefSeq and non-RefSeq sequences are emphasized during probe design and selection differs among the platforms.

[f]The number of Entrez genes specifically assayed, through any of their alternative transcript variants, is given in these columns. Probes with a few gaps or mismatches were permitted, but at the same time, probes with even a minor risk of cross-hybridization to another gene (with up to 30% mismatches) were ignored.

[g]Genes, not yet in Entrez, are supported by cDNAs in GenBank, and are described in AceView. The sum of genes in Entrez (via AceView) and genes not in Entrez that mapped to probes (via AceView) is the total number of genes in the AceView database that are matched by each platform under the mapping criteria chosen for this study.

[h]For the two PCR-based platforms (GEX and TAQ), no exact sequence mapping was conducted. Consequently, assay annotation information provided by the manufacturers was used to determine cross-platform mapping.

[i]The union and intersection numbers are based on the six high-density microarray platforms (ABI, AFX, AG1, GEH, ILM, and NCI).

F. Number of Genes Used in Analyses

Table S6. Number of Genes Used in Within-Platform Analyses

Microarray Platform	MAQC Sample	Subset of Generally Detected Genes (*i.e.*, Detected ≥ 3 Replicates)				Subset of Perfectly Detected Genes (*i.e.*, Detected in All Replicates)[c]			
		Site 1[a]	Site 2[a]	Site 3[a]	All Sites[b]	Site 1	Site 2	Site 3	All Sites
Applied Biosystems	A	8984	8944	9703	8615	8097	8300	8875	7732
	B	9055	8932	9645	8632	8249	7828	8556	7484
	C	9549	9357	9777	9081	8749	8301	8846	7999
	D	9510	9300	9585	8980	8696	8195	8978	7926
Affymetrix	A	8565	8270	8542	8016	7966	7671	8028	7389
	B	8391	8023	8394	7777	7757	7328	7780	7052
	C	8883	8608	8944	8407	8289	7918	8402	7725
	D	8919	8529	8785	8333	8259	7901	8112	7643
Agilent (one color)	A	9318	9375	9940	9164	9060	9118	9540	8810
	B	9153	9280	9827	8972	8673	8861	9543	8504
	C	9664	9815	10324	9561	9329	9429	9849	9067
	D	9676	9819	10322	9544	9141	9553	9894	8980
GE Healthcare	A	10869	10982	10827	10619	10036	10462	10370	9475
	B	10897	10927	10759	10593	10293	10433	10281	9652
	C	11085	11171	11031	10847	10559	10576	10537	9811
	D	11130	11149	11055	10861	10508	10310	10664	9605
Illumina	A	8872	8709	8505	8361	8568	8241	8062	7952
	B	8966	8779	8522	8407	8628	8301	8085	7961
	C	9137	8936	8832	8683	8939	8520	8421	8277
	D	9339	9069	8854	8756	9013	8591	8479	8353
NCI_Operon	A	12029	11527	N/A	N/A	11987	11852	N/A	N/A
	B	12030	11671			11958	11782		
	C	12024	11522			11992	11846		
	D	12050	11648			11995	11939		
Eppendorf[d]	A	160	171	174	157	N/A	N/A	N/A	N/A
	B	93	119	159	90				
	C	131	170	181	130				
	D	128	151	205	128				

[a]Numbers used in intra-site repeatability analysis, including replicate CV distribution (**Figure 1**) and the average of the three replicate CV medians (**Figure 2**).
[b]Numbers used in inter-site reproducibility analysis, including total CV median (**Figure 2**).
[c]Numbers used in concordant detection calls (**Figure 3**).
[d]Eppendorf platform is an open system. Different scanner technologies and settings have been used by the different test sites.

Table S7. Size of Differentially Expressed Gene Lists

Platform	Test Site	Gene List Size[a]
Applied Biosystems	ABI_1	4704
	ABI_2	4206
	ABI_3	4648
Affymetrix	AFX_1	4441
	AFX_2	4232
	AFX_3	4541
Agilent (one color)	AG1_1	5383
	AG1_2	5458
	AG1_3	5249
GE Healthcare	GEH_1	4228
	GEH_2	4685
	GEH_3	4763
Illumina	ILM_1	4153
	ILM_2	4248
	ILM_3	3763
NCI_Operon	NCI_1	3158
	NCI_2	2097

[a]Number of genes from the 12,091 common set with fold change > 2 and P value < 0.001, which are used in analysis of Gene List Agreement (**Figure 4**).

Table S8. Number of Genes in Log Ratio Studies

Test Site	ABI_1	ABI_2	ABI_3	AFX_1	AFX_2	AFX_3	AG1_1	AG1_2	AG1_3	EPP_1	EPP_2	EPP_3	GEH_1	GEH_2	GEH_3	ILM_1	ILM_2	ILM_3	NCI_1	NCI_2	GEX_1	QGN_1	TAQ_1
ABI_1		7679	7999	6544	6182	6481	7046	7171	7417	110	123	151	7677	7719	7646	6912	6797	6648	8051	7901	118	116	528
ABI_2	7679		7931	6570	6205	6503	7017	7133	7370	112	124	152	7576	7614	7545	6883	6778	6628	7935	7799	120	117	523
ABI_3	7999	7931		6920	6535	6858	7493	7623	7937	116	129	157	8293	8344	8251	7320	7186	7013	8786	8599	125	125	567
AFX_1	6544	6570	6920		6821	7137	6786	6910	7064	107	116	138	7182	7198	7153	6841	6734	6584	7388	7295	119	103	469
AFX_2	6182	6205	6535	6821		6972	6385	6500	6662	103	112	133	6814	6821	6774	6441	6345	6196	7034	6943	107	100	451
AFX_3	6481	6503	6858	7137	6972		6706	6835	7015	107	117	139	7175	7189	7142	6771	6658	6502	7418	7315	115	104	472
AG1_1	7046	7017	7493	6786	6385	6706		8178	8335	120	130	159	7941	7994	7907	7202	7078	6944	8300	8170	122	116	532
AG1_2	7171	7133	7623	6910	6500	6835	8178		8419	117	127	157	8079	8133	8056	7325	7193	7034	8433	8301	124	114	547
AG1_3	7417	7370	7937	7064	6662	7015	8335	8419		122	134	167	8581	8649	8552	7518	7372	7192	9089	8911	131	121	595
EPP_1	110	112	116	107	103	107	120	117	122		122	127	119	120	118	113	112	111	125	125	25	36	53
EPP_2	123	124	129	116	112	117	130	127	134	122		146	135	138	135	123	122	119	144	143	28	44	64
EPP_3	151	152	157	138	133	139	159	157	167	127	146		168	173	167	148	146	143	188	185	31	54	84
GEH_1	7677	7576	8293	7182	6814	7175	7941	8079	8581	119	135	168		10262	10210	7684	7512	7299	10389	10012	147	147	670
GEH_2	7719	7614	8344	7198	6821	7189	7994	8133	8649	120	138	173	10262		10175	7727	7550	7337	10480	10091	146	151	680
GEH_3	7646	7545	8251	7153	6774	7142	7907	8056	8552	118	135	167	10210	10175		7671	7493	7286	10256	9898	148	145	660
ILM_1	6912	6883	7320	6841	6441	6771	7202	7325	7518	113	123	148	7684	7727	7671		7764	7533	7985	7864	120	110	516
ILM_2	6797	6778	7186	6734	6345	6658	7078	7193	7372	112	122	146	7512	7550	7493	7764		7490	7793	7674	120	109	505
ILM_3	6648	6628	7013	6584	6196	6502	6944	7034	7192	111	119	143	7299	7337	7286	7533	7490		7566	7463	117	103	484
NCI_1	8051	7935	8786	7388	7034	7418	8300	8433	9089	125	144	188	10389	10480	10256	7985	7793	7566		11416	152	177	769
NCI_2	7901	7799	8599	7295	6943	7315	8170	8301	8911	125	143	185	10012	10091	9898	7864	7674	7463	11416		147	170	740
GEX_1	118	120	125	119	107	115	122	124	131	25	28	31	147	146	148	120	120	117	152	147		39	82
QGN_1	116	117	125	103	100	104	116	114	121	36	44	54	147	151	145	110	109	103	177	170	39		157
TAQ_1	528	523	567	469	451	472	532	547	595	53	64	84	670	680	660	516	505	484	769	740	82	157	

[a]Numbers used in Compression/Expansion (**Figure 5A**) and Rank Correlation (**Figure 5B**) analyses.

G. Raw Data Used in Figure Representations

Table S9. Gene List Agreement Data

Test Site X

Test Site Y	ABI_1	ABI_2	ABI_3	AFX_1	AFX_2	AFX_3	AG1_1	AG1_2	AG1_3	GEH_1	GEH_2	GEH_3	ILM_1	ILM_2	ILM_3	NCI_1	NCI_2
ABI_1		82.9	87.8	70.5	67.5	70.7	77.7	79.6	78.9	63.1	66.4	66.6	67.5	69.5	63.6	47.6	32.8
ABI_2	92.7		92.2	74.3	71.7	74.2	80.0	82.2	82.2	65.4	68.4	68.7	71.0	72.8	67.1	51.0	35.4
ABI_3	88.9	83.4		73.1	70.2	73.4	79.1	81.4	80.8	64.8	67.9	68.6	69.8	71.1	65.1	49.6	33.6
AFX_1	74.6	70.3	76.5		91.1	95.8	89.1	89.7	87.4	72.2	75.9	76.0	79.4	79.8	74.2	56.8	38.4
AFX_2	75.0	71.2	77.2	95.6		98.5	89.4	89.3	87.3	73.3	76.9	77.0	79.4	79.5	74.3	58.4	39.7
AFX_3	73.2	68.7	75.1	93.7	91.8		88.4	88.3	85.6	71.6	75.6	75.8	77.3	77.6	71.9	56.2	37.9
AG1_1	67.9	62.5	68.3	73.5	70.3	74.5		91.4	87.5	64.3	69.1	69.3	67.4	68.6	62.9	47.8	32.9
AG1_2	68.6	63.3	69.3	73.0	69.3	73.5	90.1		88.4	64.0	68.0	68.8	68.5	69.8	63.7	47.2	32.0
AG1_3	70.7	65.8	71.5	73.9	70.4	74.1	89.7	91.9		64.1	67.5	68.1	70.1	71.5	65.1	48.3	33.0
GEH_1	70.2	65.1	71.3	75.9	73.4	76.9	81.9	82.6	79.5		91.8	95.5	70.6	71.3	65.3	51.3	35.8
GEH_2	66.7	61.4	67.4	71.9	69.5	73.2	79.3	79.2	75.7	82.8		89.8	65.8	66.5	61.0	48.6	33.6
GEH_3	65.8	60.7	67.0	70.9	68.4	72.2	78.3	78.8	75.0	84.8	88.3		65.3	65.9	60.2	47.6	32.9
ILM_1	76.5	71.9	78.1	84.9	80.9	84.5	87.4	90.1	88.5	71.9	74.3	74.8		95.4	87.4	55.9	38.3
ILM_2	77.0	72.0	77.8	83.5	79.2	82.9	86.9	89.7	88.3	71.0	73.4	73.9	93.3		86.1	54.4	37.5
ILM_3	79.5	75.0	80.4	87.6	83.6	86.8	90.0	92.4	90.8	73.4	76.0	76.2	96.4	97.2		57.8	40.4
NCI_1	70.9	68.0	73.1	79.9	78.3	80.7	81.4	81.5	80.3	68.7	72.2	71.7	73.6	73.2	68.9		51.3
NCI_2	73.5	71.0	74.5	81.4	80.2	82.1	84.4	83.3	82.7	72.2	75.0	74.8	75.9	75.9	72.6	77.3	

Data presented in ***Figure 4****. Gene list agreement is defined as the percent overlap of genes on the list for Test Site Y that are also present on the list for Test Site X. The size of the list of differentially expressed genes for each test site is reported in* ***Table S7****.*

Table S10. Compression/Expansion of Log Ratio

Test Site X

Test Site Y	ABI_1	ABI_2	ABI_3	AFX_1	AFX_2	AFX_3	AG1_1	AG1_2	AG1_3	EPP_1	EPP_2	EPP_3	GEH_1	GEH_2	GEH_3	ILM_1	ILM_2	ILM_3	NCI_1	NCI_2	GEX_1	QGN_1	TAQ_1
ABI_1		-.010	-.021	-.113	-.144	-.114	.064	.090	.049	-.158	-.182	-.216	-.244	-.203	-.184	-.145	-.124	-.163	-.439	-.278	.306	.142	.304
ABI_2	.010		-.010	-.107	-.138	-.105	.078	.101	.065	-.202	-.211	-.241	-.223	-.181	-.162	-.135	-.115	-.152	-.419	-.247	.304	.145	.321
ABI_3	.022	.010		-.102	-.133	-.101	.095	.122	.081	-.136	-.152	-.166	-.219	-.176	-.155	-.133	-.112	-.153	-.425	-.262	.316	.119	.328
AFX_1	.127	.120	.114		-.024	.006	.168	.187	.161	-.033	-.112	-.085	-.129	-.098	-.075	-.041	-.020	-.065	-.292	-.108	.364	.149	.282
AFX_2	.168	.159	.154	.025		.034	.209	.228	.204	-.032	-.119	-.087	-.095	-.063	-.038	-.014	.008	-.034	-.257	-.070	.261	.180	.331
AFX_3	.129	.118	.112	-.006	-.033		.171	.191	.170	-.057	-.135	-.117	-.135	-.103	-.078	-.048	-.025	-.070	-.300	-.127	.240	.128	.251
AG1_1	-.060	-.072	-.086	-.144	-.173	-.146		.021	-.031	-.177	-.226	-.311	-.287	-.246	-.227	-.179	-.159	-.196	-.477	-.328	.098	-.054	.140
AG1_2	-.083	-.092	-.109	-.157	-.186	-.161	-.021		-.046	-.187	-.228	-.330	-.313	-.274	-.256	-.197	-.178	-.213	-.496	-.356	.076	-.146	.069
AG1_3	-.046	-.061	-.075	-.139	-.169	-.146	.032	.048		-.224	-.272	-.308	-.285	-.243	-.223	-.172	-.151	-.188	-.479	-.337	.126	-.037	.182
EPP_1	.188	.253	.157	.035	.033	.061	.215	.229	.289		.008	.012	-.026	.007	.007	.016	.038	-.013	-.221	.016	.525	.139	.562
EPP_2	.223	.267	.179	.127	.135	.156	.292	.295	.374	-.008		.024	-.019	.035	.046	.111	.137	.071	-.217	.060	.542	.378	.891
EPP_3	.275	.318	.199	.093	.096	.133	.451	.492	.445	-.012	-.024		.041	.092	.107	.088	.102	.042	-.266	-.013	.525	.283	.911
GEH_1	.323	.288	.280	.148	.105	.156	.403	.456	.399	.027	.019	-.040		.038	.075	.111	.140	.080	-.265	-.072	1.044	.365	.774
GEH_2	.255	.221	.213	.108	.067	.115	.326	.377	.320	-.007	-.034	-.084	-.037		.037	.063	.091	.035	-.304	-.125	.973	.284	.665
GEH_3	.225	.193	.184	.081	.040	.085	.294	.343	.287	-.007	-.044	-.097	-.070	-.036		.036	.063	.006	-.337	-.161	.855	.243	.631
ILM_1	.169	.156	.154	.043	.014	.050	.219	.245	.208	-.016	-.100	-.081	-.100	-.059	-.034		.019	-.032	-.305	-.120	.455	.259	.401
ILM_2	.141	.131	.127	.020	-.008	.026	.189	.217	.178	-.036	-.120	-.093	-.122	-.083	-.059	-.019		-.048	-.317	-.133	.351	.215	.382
ILM_3	.195	.179	.180	.069	.035	.075	.244	.271	.232	.013	-.066	-.041	-.074	-.034	-.006	.033	.050		-.280	-.082	.403	.268	.448
NCI_1	.784	.721	.739	.412	.346	.428	.912	.985	.921	.284	.277	.363	.361	.437	.509	.438	.464	.389		.217	1.481	1.386	1.672
NCI_2	.385	.327	.355	.121	.076	.145	.488	.553	.507	-.016	-.057	.013	.078	.142	.191	.137	.153	.089	-.178		.915	.780	1.213
GEX_1	-.234	-.233	-.240	-.267	-.207	-.193	-.089	-.070	-.112	-.344	-.351	-.344	-.511	-.493	-.461	-.313	-.260	-.287	-.597	-.478		-.284	-.012
QGN_1	-.124	-.127	-.107	-.129	-.153	-.113	.057	.171	.038	-.122	-.274	-.220	-.267	-.221	-.195	-.206	-.177	-.211	-.581	-.438	.397		.286
TAQ_1	-.233	-.243	-.247	-.220	-.249	-.201	-.123	-.065	-.154	-.360	-.471	-.477	-.436	-.399	-.387	-.286	-.277	-.310	-.626	-.548	.012	-.222	

*Data represented in **Figure 5A**. Compression/expansion is defined as the percent difference from equivalency between platform/sites (corresponding to a slope value 1 for the best fitted line using orthogonal regression) of the log ratio differential expression using A and B replicates.*

Table S11. Rank Correlation of Log Ratio

Test Site X

Test Site Y	ABI_1	ABI_2	ABI_3	AFX_1	AFX_2	AFX_3	AG1_1	AG1_2	AG1_3	EPP_1	EPP_2	EPP_3	GEH_1	GEH_2	GEH_3	ILM_1	ILM_2	ILM_3	NCI_1	NCI_2	GEX_1	QGN_1	TAQ_1
ABI_1		.985	.982	.871	.868	.870	.875	.877	.872	.848	.834	.835	.824	.834	.815	.870	.873	.867	.742	.725	.757	.836	.847
ABI_2	.985		.984	.875	.873	.875	.873	.874	.871	.850	.845	.839	.829	.835	.819	.874	.877	.870	.745	.725	.742	.860	.843
ABI_3	.982	.984		.882	.881	.882	.878	.880	.874	.864	.853	.842	.825	.833	.817	.881	.882	.873	.741	.722	.739	.842	.842
AFX_1	.871	.875	.882		.994	.996	.936	.936	.933	.855	.846	.847	.872	.872	.856	.941	.941	.936	.831	.808	.814	.864	.905
AFX_2	.868	.873	.881	.994		.997	.931	.931	.929	.834	.818	.830	.870	.871	.854	.935	.935	.929	.833	.808	.796	.845	.903
AFX_3	.870	.875	.882	.996	.997		.935	.935	.932	.861	.853	.855	.871	.871	.857	.939	.938	.932	.832	.808	.812	.869	.907
AG1_1	.875	.873	.878	.936	.931	.935		.986	.979	.901	.869	.867	.859	.864	.851	.930	.930	.927	.781	.765	.842	.876	.887
AG1_2	.877	.874	.880	.936	.931	.935	.986		.980	.903	.861	.872	.861	.865	.853	.932	.931	.927	.779	.763	.852	.863	.902
AG1_3	.872	.871	.874	.933	.929	.932	.979	.980		.911	.878	.870	.845	.849	.834	.926	.926	.922	.772	.757	.831	.858	.892
EPP_1	.848	.850	.864	.855	.834	.861	.901	.903	.911		.954	.964	.821	.817	.826	.911	.906	.903	.815	.796	.785	.904	.908
EPP_2	.834	.845	.853	.846	.818	.853	.869	.861	.878	.954		.980	.779	.772	.789	.904	.903	.893	.762	.729	.698	.803	.798
EPP_3	.835	.839	.842	.847	.830	.855	.867	.872	.870	.964	.980		.790	.796	.812	.894	.895	.882	.748	.720	.590	.834	.839
GEH_1	.824	.829	.825	.872	.870	.871	.859	.861	.845	.821	.779	.790		.968	.980	.863	.861	.862	.723	.709	.731	.805	.840
GEH_2	.834	.835	.833	.872	.871	.871	.864	.865	.849	.817	.772	.796	.968		.972	.865	.865	.864	.727	.710	.750	.800	.855
GEH_3	.815	.819	.817	.856	.854	.857	.851	.853	.834	.826	.789	.812	.980	.972		.853	.851	.851	.705	.690	.730	.770	.823
ILM_1	.870	.874	.881	.941	.935	.939	.930	.932	.926	.911	.904	.894	.863	.865	.853		.994	.990	.801	.781	.834	.880	.906
ILM_2	.873	.877	.882	.941	.935	.938	.930	.931	.926	.906	.903	.895	.861	.865	.851	.994		.991	.802	.780	.826	.906	.904
ILM_3	.867	.870	.873	.936	.929	.932	.927	.927	.922	.903	.893	.882	.862	.864	.851	.990	.991		.802	.785	.832	.872	.897
NCI_1	.742	.745	.741	.831	.833	.832	.781	.779	.772	.815	.762	.748	.723	.727	.705	.801	.802	.802		.901	.755	.821	.851
NCI_2	.725	.725	.722	.808	.808	.808	.765	.763	.757	.796	.729	.720	.709	.710	.690	.781	.780	.785	.901		.792	.775	.830
GEX_1	.757	.742	.739	.814	.796	.812	.842	.852	.831	.785	.698	.590	.731	.750	.730	.834	.826	.832	.755	.792		.834	.865
QGN_1	.836	.860	.842	.864	.845	.869	.876	.863	.858	.904	.803	.834	.805	.800	.770	.880	.906	.872	.821	.775	.834		.902
TAQ_1	.847	.843	.842	.905	.903	.907	.887	.902	.892	.908	.798	.839	.840	.855	.823	.906	.904	.897	.851	.830	.865	.902	

*Data represented in **Figure 5B**. The Spearman rank correlations of the log ratios compare the relative position of a gene in the Test Site X rank order against its position in the Test Site Y rank order.*

H. Additional Analyses from MAQC Study

Table S12. Average Spearman Rank Correlation (*r*) of Log (Sample B/A)

Platform on Y Axis	Platform on X Axis									
	ABI	**AFX**	**AG1**	**EPP**	**GEH**	**ILM**	**NCI**	**GEX**	**QGN**	**TAQ**
ABI	0.984	*6533*	*7356*	*130*	*7852*	*6907*	*8179*	*121*	*119*	*539*
AFX	0.875	0.995	*6763*	*119*	*7050*	*6564*	*7232*	*114*	*102*	*464*
AG1	0.875	0.933	0.981	*137*	*8210*	*7206*	*8534*	*126*	*117*	*558*
EPP	0.845	0.844	0.881	0.966	*141*	*126*	*152*	*28*	*45*	*67*
GEH	0.826	0.866	0.853	0.800	0.973	*7507*	*10188*	*147*	*148*	*670*
ILM	0.874	0.936	0.928	0.899	0.860	0.992	*7724*	*119*	*107*	*502*
NCI	0.733	0.820	0.770	0.762	0.711	0.792	0.901	*150*	*174*	*755*
GEX	0.746	0.807	0.842	0.691	0.737	0.831	0.773	*NA*	*17*	*82*
QGN	0.846	0.859	0.865	0.847	0.792	0.886	0.798	0.834	*NA*	*157*
TAQ	0.844	0.905	0.894	0.848	0.839	0.902	0.841	0.865	0.902	*NA*

The table displays summaries of both inter-site reproducibility, when the platform is compared to itself, or between-platform comparability when one platform is compared to a different platform. The platforms are labeled according to the codes presented in Table 1. **Bottom left**) The average over paired sites for the indicated platform pairings of Spearman rank correlation (r) for the log (B/A) values between each paired site-platform **Top right**) Italicized numbers highlighted in grey indicate how many genes were detected on average in at least three of the five replicate A and B (both) samples in both paired site-platforms.

Table S13. Average Slope (β) of the Fitted Line

Platform on Y Axis	Platform on X Axis									
	ABI	**AFX**	**AG1**	**EPP**	**GEH**	**ILM**	**NCI**	**GEX**	**QGN**	**TAQ**
ABI	1.00	1.13	0.92	1.23	1.24	1.16	1.55	0.76	0.88	0.76
AFX	0.88	1.00	0.84	1.10	1.10	1.04	1.25	0.78	0.87	0.78
AG1	1.08	1.19	1.00	1.34	1.36	1.22	1.73	0.91	1.09	0.89
EPP	0.82	0.91	0.75	1.00	0.97	0.94	1.14	0.65	0.79	0.56
GEH	0.81	0.91	0.74	1.03	1.00	0.94	1.29	0.51	0.77	0.59
ILM	0.86	0.97	0.82	1.07	1.07	1.00	1.28	0.71	0.80	0.71
NCI	0.65	0.81	0.59	0.89	0.79	0.79	1.02	0.46	0.49	0.41
GEX	1.31	1.29	1.10	1.53	1.96	1.40	2.20	NA	1.40	1.01
QGN	1.14	1.15	0.92	1.27	1.30	1.25	2.08	0.72	NA	0.78
TAQ	1.32	1.29	1.13	1.79	1.69	1.41	2.44	0.99	1.29	NA

We fitted orthogonal regression lines of log Ratio (B/A) values between each possible pairing of site-platform (eg, AG1_1 vs ILM_3) using only those transcripts detected (>=3 replicates) in each site-platform, and averaged the estimates of slope from the regression results for the platform pairings indicated. If the value in a cell is less than 1, then the signal of detected transcripts for the platform indicated in the column (Platform on Y-axis) is generally compressed versus the platform indicated in the row (Platform on X-axis). One can see that almost all microarray platforms are generally compressed relative to the alternative platforms, although there are exceptions.

I. Apparent Power Analysis

We used a graphical method based on a modification of the two-sided t-test power analysis in an effort to illustrate the effect of the within-group variation caused by differences in inter-site reproducibility and intrinsic platform-dependent factors, such as the effect size between two groups for every probe on the microarray. The tool is useful for quality assessment, but the results should not be confused with previous microarray power analysis methods where all the parameters are designated with the exception of standard deviation [Hwang, D., *et al. Bioinformatics* **18**, 1184-1193, 2002; Page, G.P. *et al. BMC Bioinformatics* **7**, 84, 2006; Seo, J., *et al*. *Bioinformatics* **22**, 808-814, 2006; Tibshirani, R. *BMC Bioinformatics* **7**, 106, 2006].

Using our modification of the power analysis, apparent power was calculated for two separate group comparisons at each site: sample A replicates vs. sample B replicates as well as sample C replicates vs. sample D replicates. For our analysis, we used the measured average difference between groups (*i.e.,* A vs. B replicates, or C vs. D replicates) and calculated the estimated pooled standard deviation (s_{pooled}) for each gene based on the signal intensities generated for each of the two experimental groups being compared for each site. The results are expressed as the percentage of genes on the y-axis with a calculated power equal to or greater than a given power on the x-axis, shown in the figure below.

For each comparison, the power analyses for all test sites using the same microarray platform are grouped to display the extremes of test site performance. As expected, the comparisons of the A vs. B replicates demonstrated greater average power than the comparisons of the C vs. D replicates, because the titrated samples can show at most a 3-fold change in gene abundance. Cumulatively each platform has similar power across the 12,091 set of common genes, but for each platform there was at least one site that showed a substantial loss of power due to increased technical noise. For example, Applied Biosystems test site 2 had a lower labeling efficiency in sample type A (Tong, W. *et al. Nat Biotechnol* **24**(9), 2006) which impacted its performance in the power analysis relative to the other platforms for the A vs. B comparison. An increase in power was observed at Illumina site 1 compared to sites 2 and 3. This site 1 also consistently had the lowest CV distribution (**Fig. 1**) while not being any more compressed in signal (**Fig. 5a**) than the other two Illumina sites These relative differences illustrate the importance of a detailed review of laboratory performance in microarray facilities.

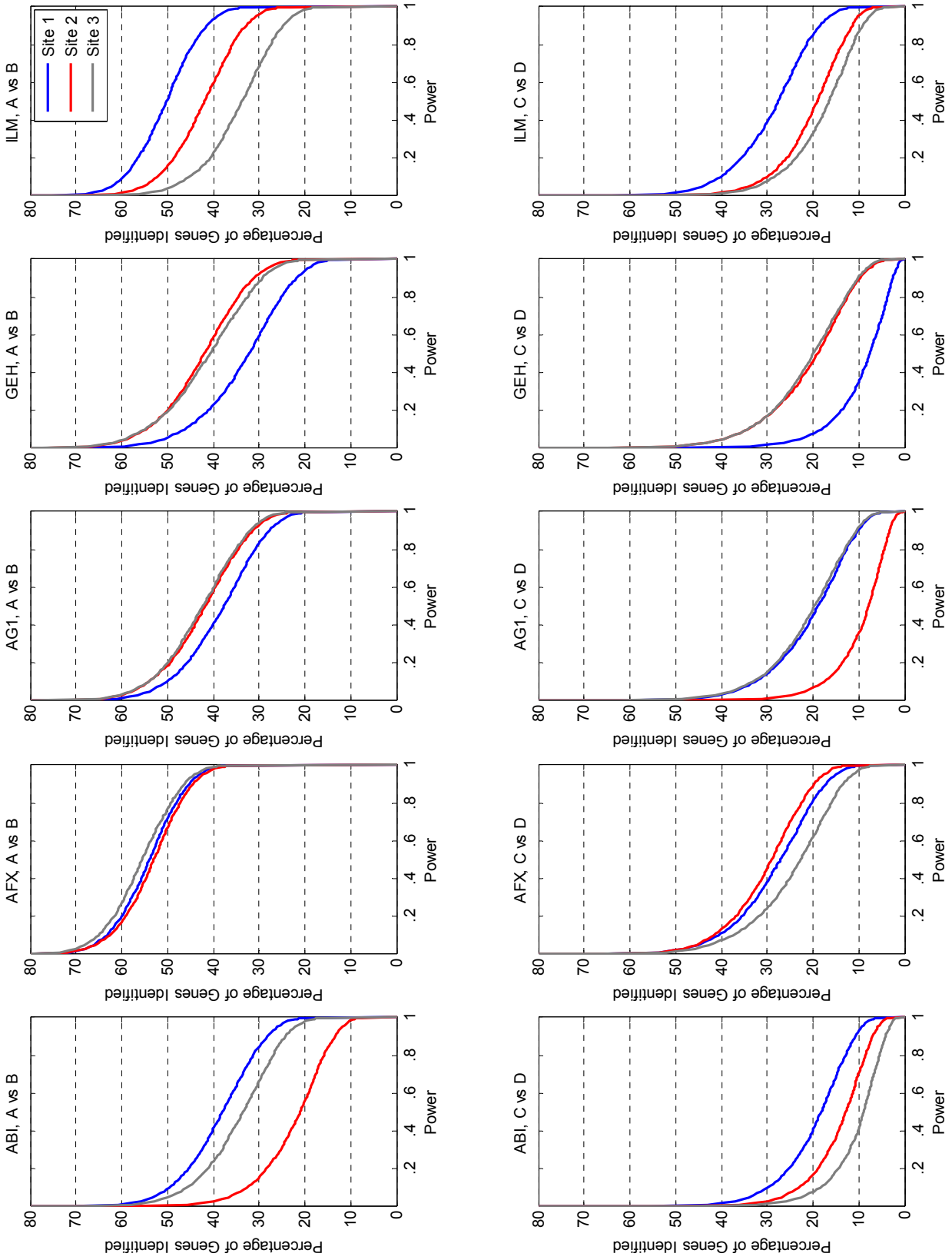
Apparent Power Analysis Methods. The power analysis is based on data from the 12,091 common genes set. No filtering related to gene detection was performed. Three probes were removed from the analysis because data were missing in three or more of the groups. Microarray data were normalized using the manufacturer's suggested method and $\log_2$ transformed. After $\log_2$ transformation, the signal for all microarrays approximated a normal distribution (data not shown). We implemented a novel power analysis based on Warnes & Liu's method (www.bioconductor.org/repository/devel/vignette/ssize.pdf) with four key modifications: 1) the average difference between groups was explicitly calculated for each probe; 2) a pooled estimate of σ (s_{pooled}) was used; 3) experimentally derived power was plotted; and 4) the method was generalized so it could be used for all microarray platforms. The key component of this analysis is the generation of a cumulative plot of the proportion of genes achieving a desired power for a given sample size ($n = 5$), multiple test corrected α ($0.05/n$) using the Bonferonni method, and a probe-by-probe s_{pooled} for each site and measure difference between groups at each site.

Table S14. Number of Genes with Power ≥ 0.8

Microarray Platform	A vs. B Comparison			C vs. D Comparison		
	Site 1	Site 2	Site 3	Site 1	Site 2	Site 3
Applied Biosystems	3782	1916	3261	1560	1031	666
Affymetrix	5821	5710	5959	2448	2731	1931
Agilent (one color)	3733	4184	4239	1535	531	1578
GE Healthcare	3044	4192	3934	455	1521	1553
Illumina	5367	4339	3331	2593	1656	1359

Figure S1. Apparent Power Analysis.

The power analysis used the standard formula for the power of two-group t-test with pooled σ. The mean difference between groups of sample type A replicates vs. sample type B replicates (**top**) as well as sample type C replicates vs. sample type D replicates (**bottom**) were computed for each gene and the σ at each location was calculated independently and plotted by location. The x-axis is the calculated power and the y-axis is the percent of genes that have that power or greater. The number of genes with power ≥ 0.8 for each test site is reported in Supplemental **Table S14**. Only genes from the 12,091 common set were included in the analysis. Results for each platform are displayed in separate plots per comparison. The power lines for each test site are colored as indicated. As described in the text, data from some platforms were omitted from these calculations due to quality issues. The platforms and sample types are labeled according to the nomenclature presented in **Table 1**.


ILM, A vs B
ILM, C vs D
GEH, A vs B
GEH, C vs D
AG1, A vs B
AG1, C vs D
AFX, A vs B
AFX, C vs D
ABI, A vs B
ABI, C vs D
Site 1
Site 2
Site 3
Percentage of Genes Identified
Power
0
10
20
30
40
50
60
70
80
.2
.4
.6
.8
1

J. Reproducibility of Lists of Differentially Expressed Genes – POG Results

Gene Ranking (Selection) Rules: Six gene ranking (selection) methods were examined: (1) Fold-change ranking; (2) Fold-change with *P* cutoff of 0.05; (3) Fold-change ranking with *P* cutoff of 0.01; (4) *P* ranking; (5) *P* ranking with fold-change cutoff of 2.0; (6) *P* ranking with fold-change cutoff of 1.4. When a cutoff value (*e.g*., $P<0.05$) is imposed with a particular ranking metric (*e.g*., fold-change), frequently the lists of candidate genes that meet the cutoff value will not be the same for the two test sites as a result of differences in inter-site variations. Such differences are part of the gene selection process and have been carried over to the gene ranking/selection stage. *P* values are derived from simple t-tests using log2 intensity data assuming equal variance.

POG (Percentage of Overlapping Genes): The POG (percentage of overlapping genes) is a measure of the reproducibility of lists of differentially expressed genes (Shi, L. *et al.* Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics* **6**(Suppl 2), S12, 2005). The POG graph (**Fig. S2**) can be viewed as a combination of many Venn diagrams based on various thresholds of fold-change and/or *P* which are arbitrary. The number of genes considered as differentially expressed is denoted as 2L, where L is both the number of genes up- and down-regulated. The number of genes available for ranking and selection in one direction, L, varies from 1 to 6000 (with a step of one) or when there are no more genes in one regulation direction, corresponding to 2L varying from 2 to 12,000. Directionality of gene regulation is considered in POG calculations; genes selected by two test sites but with different regulation directionalities are considered as discordant. The formula for calculating POG is:

$$POG = 100*(DD+UU)/2L$$

where DD and UU are the number of commonly down- or up-regulated genes, respectively, from the two lists, and L is the number of genes selected from the up- or down-regulation directionality. To overcome the confusion of different numbers for the denominator, in our POG calculations we deliberately selected an equal number of up-regulated and down-regulated genes, L.

Figure S2 shows that the reproducibility of differentially expressed genes is dramatically impacted by several factors: gene ranking/selection methods, the inherent differences between the samples being compared, and the choice of thresholds (corresponding to the number of genes selected as differentially expressed). Given the same data set, the reproducibility of microarray results can be dramatically impacted by the choice of different gene selection methods. Data from Affymetrix's test sites were used to create **Figure S2**, but similar results have observed in inter-site comparisons of data from other platforms or in cross-platform comparisons. Furthermore, similar results were obtained when a rat toxicogenomics data set was analyzed (Guo, L. *et al. Nat. Biotechnol.* **24**(9), 2006).

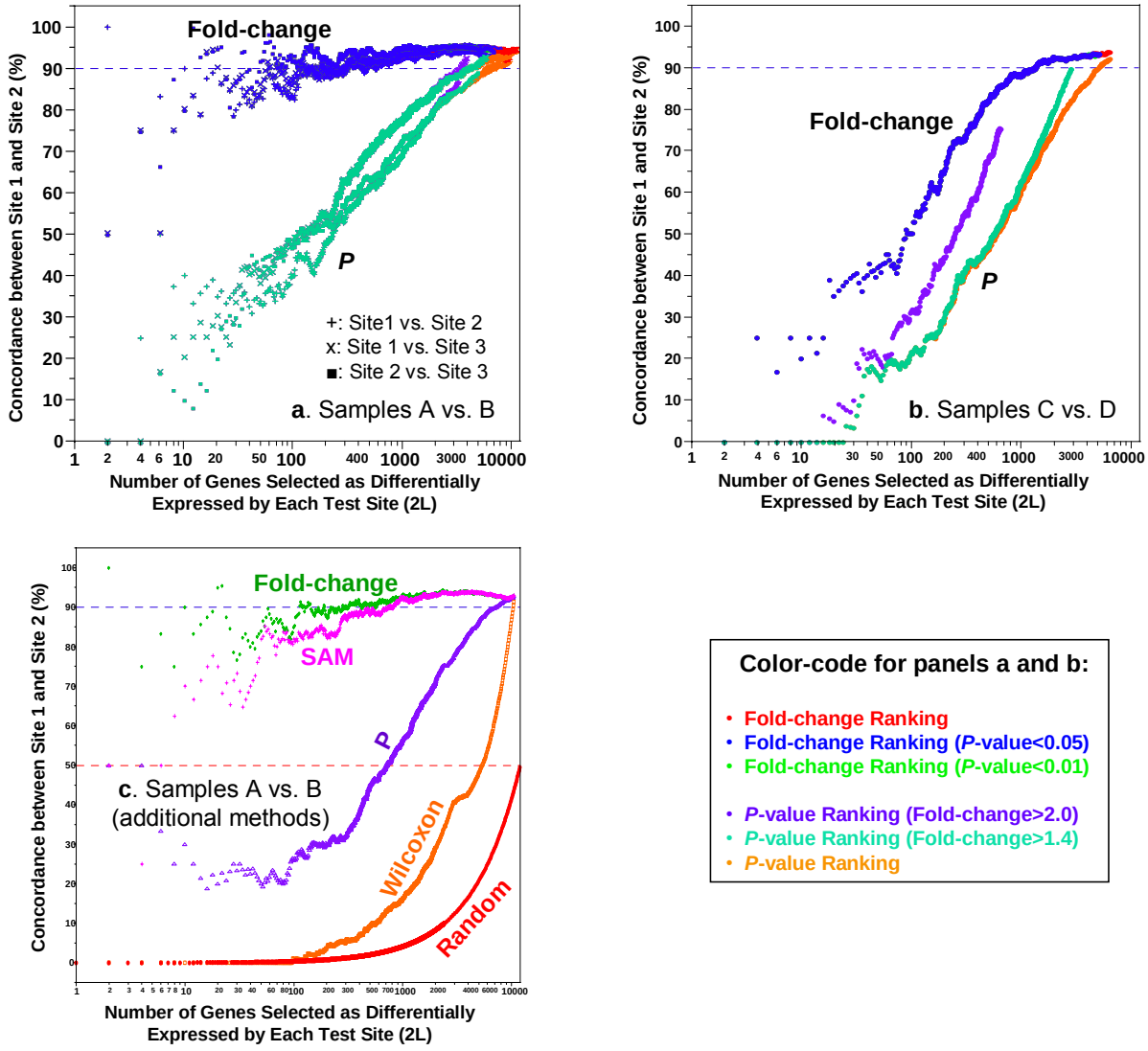


Figure S2. Reproducibility of lists of differentially expressed genes – POG results. (a) Sample A versus Sample B (Site 1 vs. Site 2, Site 1 vs. Site 3, and Site 2 vs. Site 3); (b) Sample C versus Sample D (Site 1 vs. Site 2); and (c) Sample C versus Sample D with additional methods compared (Site 1 vs. Site 2). Each panel represents the concordance results (POG) for comparing Affymetrix's test sites. The x-axis represents the number of genes selected as differentially expressed (corresponding to different thresholds), and the y-axis is the percentage (%) of genes common to the two gene lists derived from two test sites at a given number of selected genes. For the Wilcoxon rank-sum tests (panel c), there were many ties, *i.e*., many genes exhibited the same level of statistical significance because of the small sample sizes (five replicates for each group). The tied genes from each test site were broken (ranked) by random ordering. Concordance between genes selected completely at random is shown in red and reaches only 50% when all candidate genes are declared as differentially expressed due to directionality disagreement. The POG results by SAM (pink line), although greatly improved over that of simple t-test statistic (purple line), approached, but did not exceed, the level of POG based on fold-change ranking (green line).