

MAQC1 samples: 100% Brain total RNA, 100% Universal RNA from Stratagene, 25% Universal-75% Brain, and 75% Universal-25% Brain.

Total RNA Titration Data

For the human total RNA titration experiment, five cRNA samples were hybridized on a single Human-6v1 BeadChip. The five samples were placed on arrays A (highest concentration) through E (lowest concentration). The box plots for the average signal intensities for all the arrays show that the intensities diminish from array A to array E (Figure 1). Although both average normalization and rank invariant normalization scale the data, average normalization scales all the arrays tighter around the mean of the arrays. Cubic spline normalization rescales all arrays so that the means are equal.

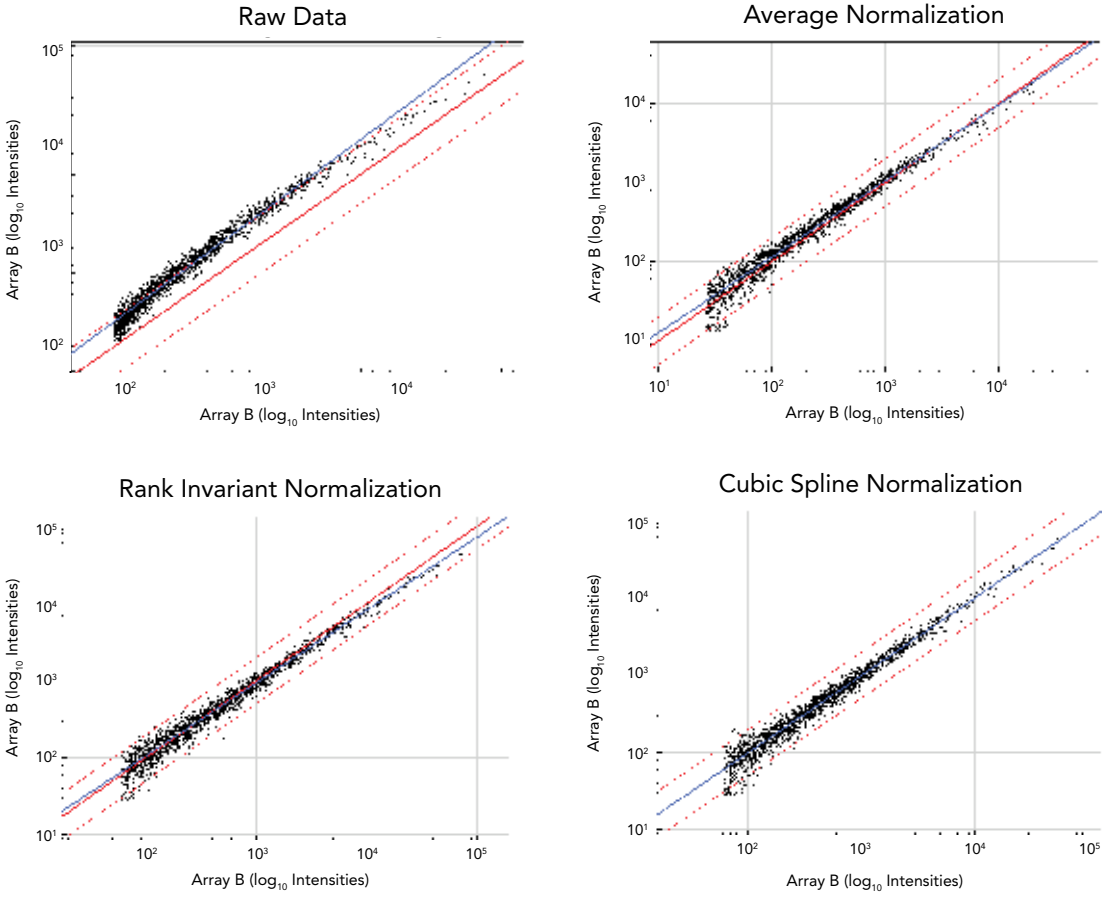
A log scatter plot displaying the intensity signals from array A versus array B demonstrates the nonlinear relationship of the raw data and reveals a significant curvature in the data set. Additionally, the raw data set is displaced from the identity line, and the best fit line is offset and skewed from the identity line. (Figure 2). When adjusting nonlinear

data, cubic spline normalization achieves the greatest improvement in symmetry, and the best agreement between the identity and best fit line.

MAQC Titration Data

The MAQC data were analyzed to demonstrate the results of Bead-Studio normalization on biologically relevant samples. The samples were assayed on four Human-6v1 BeadChips. In practice, when the data are of high quality, different normalization methods do not result in drastically different analysis outcomes. The robustness of high-quality data is illustrated with box plots, scatter plots, clustering, and differential gene expression analysis of the MAQC sample profiles (Figures 3, 4, and 5, and Table 1). The box plots show considerable similarity between average normalization and cubic spline normalization in terms of scaling the median and generating a uniform interquartile range. In comparison, a larger degree of variability remains after rank invariant normalization. Scatter plots of the MAQC data also reveal only minor differences in the results of the three normalization algorithms, with

Figure 2: Effect of Average, Rank Invariant, and Cubic Spline Normalization on Nonlinear Data



The data were filtered by the detection p-value < 0.01. The identity line and two-fold boundaries are denoted in red. The best fit line is blue.

AAAGAATGATAACAGTAAACACACCTTCTGTTAACCCTTAAGATTACTTATCCACTGATTCAACGTACCCTAAACGAGTATCAATTGAGACTAAATATTAACGTACCCTAAAGAGTACCGTCTTCTGTTAACCTTAAGATTACTTATCCACTGATTCA
AATCAACGTAACGTAACGACGTATCATTAAGATTACTTATCCACTGATTCAACGTACCCTAAACGAGTATCAATTGAGACTAAATATTAACGTACCCTAAAGAGTACCGTCTTCTGTTAACCTTAAGATTACTTATCCACTGATTCAACG
AACGACGAAAAGAATGATAACAGTAAACACACCTTCTGTTAACCCTTAAGATTACTTATCCACTGATTCAACGTACCCTAAAGAGTATCAATTGAGACTAAATATTAACGTACCCTAAAGAGTACCGTCTTCTGTTAACCTTAAGATTACTTATCCACT
TTAAAGTACCATTAAAGAGTACCGTCTTCTGTTAACCCTTAAGATTACTTATCCACTGATTCAACGTACCCTAAAGAGTATCAATTGAGACTAAATATTAACGTACCCTAAAGAGTACCGTCTTCTGTTAACCTTAAGATTACTTATCCACT
AAAGAATGATAACAGTAAACACACCTTCTGTTAACCCTTAAGATTACTTATCCACTGATTCAACGTACCCTAAAGAGTATCAATTGAGACTAAATATTAACGTACCCTAAAGAGTACCGTCTTCTGTTAACCTTAAGATTACTTATCCACTGATTCA
AAAGATTACTTATCCACTGATTCAACGTTAAGATTACTTATCCACTGATTCAACGTACCCTAAAGAGTATCAATTGAGACTTCTGTTAACCCTTAAGATTACTTATCCACTGATTCAACGTACCCTAAAGAGTATCAATTGAGACTTACCAACGA
AACGTATCAATTGAGACTAAATATTAACGTACCCTAAAGAGTCTGTTAACCCTTAAGATTACTTATCCACTGATTCAACGTACCCTAAAGAGTATCAATTGAGACTAAATATTAACGTACCCTAAAGAGTACCGTCTTCTGTTAACCTTAAGATTACTTATCCACTGATTCAACG

