a) log2(PM/MM) Histograms by log2(PMxMM)



Figure 1: a) Histograms of log ratio $\log_2(PM/MM)$, stratified by quantiles of abundance, $\log_2 \sqrt{PM \times MM}$, with gray scale representing height of histogram (light grays are high and dark grays are low) for one array from the mouse data set. The histograms have been scaled so that the mode of each histogram is represented with the same gray scale. b) Histogram of log ratios for first quartile of abundance with the histogram for the defective probes represented by a darker gray. c) Like b) for abundance values between first and third quartile. d) Like b) for abundance values in the last quartile excluding the highest 5 percent. e) Like b) for the highest 5 percent of abundance.



Figure 2: Box-plots of $\log_2(PM)$ and PM - MM for the 30 arrays from data set C. Becasue PM - MM values are usually between -2000 and 10000, a reduced range is used to get a better view of the interquartile range. The bottom row are the after quantile normalization box-plots. The y-axis scale can be deduced from the plot titles.



Figure 3: MVA plots (described in text) of $\log_2(PM)$ and $\log_2(PM - MM)$ for two arrays in which the BioDn-3 gene has been spiked at 5 pM and 2 pM respectively. The large points represent the 20 spiked-in probes and the small black dots represent a random sample of non-spiked-in probes. a) and c) are before normalization and b) and d) are after quantile normalization.



Figure 4: *PM*, *MM*, *PM/MM*, and *PM* – *MM* values for each of the 20 probes representing BioB-5 (with the exception of CreX-3, all other spike-in genes behaved similarly to BioB-5) in the 12 spiked-in arrays from the varying concentration experiment plotted against concentration. The different probes are represented by the different line types and symbols. The horizontal line represents the median of the 20 BioB-5 probes for the non-spiked-in array. The dashed lines are the 25th and 75th quantiles.



Figure 5: Histograms of $\log_2(MM)$ for a array in which no probe-set was spiked along with the 3 arrays in which BioB-5 was spiked-in at concentrations of 0.5, 0.75, and 1 pM. The observed *PM* values for the 20 probes associated with BioB-5 are marked with crosses and the average with an arrow. The black curve represents the log normal distribution obtained from left-of-the-mode data.



Figure 6: Data set C box-plots. a) Averages over replicates for each gene in b) Loess curves fitted to standard deviation vs. average expression scatter-plots.



Figure 7: MVA (described in text) and qq-plots indicating the positions of differentially expressed genes ranked by their absolute log relative expression values.

SD vs. Avg for PM







Figure 8: Standard deviations (SDs) plotted against averages from 5 MGU74A mouse arrays for a random sample of 2000 defective probe sets for PM and $\log 2(PM)$. The curves are loss fits.



Figure 9: Box-plots for probe intensities $\log_2(PM)$ (green) and $\log_2(MM)$ (red), log ratios $\log_2(PM/MM)$, and differences PM - MM for 8 RGU34A rat arrays. The right column are the after quantile normalization box-plots. The y-axis scale can be deduced from the plot titles.



Figure 10: MVA plot (described in text) of $\log_2(PM)$ for all pairwise comparisons of the 5 MGU74A mouse arrays.



Figure 11: MVA plot (described in text) of $\log_2(PM)$ for all pairwise comparisons of 5 arrays, after quantile normalization.



Figure 12: MVA plots (described in text) for the 4 different expression values before (left column) and after (right column) probe-level normalization for the arrays described in Figure 6. The red large dot is from the differentially expressed probe set.

a) Histogram of log(MM) and estimated BG density



b) Quantile-quantile plot



Figure 13: a) Histograms of $\log_2(MM)$ for the array with spike-in concentration of 12.5 pM in the varying concentration series. b) Quantile-quantile plot of the *MM*s to the left of the mode of the histogram in a) compared to a log-normal distribution with mean and SD estimated from data.



Figure 14: Summary measures of expression for 10 cRNA controls and their average, over 12 concentrations denoted with black points. Yellow cloud represents the IQR of the values obtained for the rest of the probe sets. The red solid line is the median value of the expressions obtained for the 10 control genes in the array with no spike-in. The dotted red lines represent the third and first quartiles. The blue line represents a regression line fitted to the averages.



a) Model estimate compared to observed variance

Figure 15: a) Box-plot of $\log(\hat{\sigma}/SD)$ for Li and Wong's MBEI measure and AvLog(PM-BG) for each concentration in the dilution study. b) $\log(\hat{\sigma}/SD)$ plotted against log expression of Li and Wong's MBEI measure. c) $\log(\hat{\sigma}/SD)$ plotted against AvLog(PM-BG).



Figure 16: Observed ratios against true ratios for 23 combinations of spike-in triplicates. A regression line is fitted to these values for each expression measure. The horizontal lines represent the IQR of the observed ratios of the non-spiked-in probe sets. The dotted vertical lines denote the largest and smallest true ratios for which the observed ratio fell within the IQR of non-spiked in data.