Bioconductor Expression Assessment Tool for Affymetrix Oligonucleotide Arrays (affycomp)

June 8, 2011

Contents

In this report only assessment using the **HGU133** spike-In experiment are presetned. Figures 2,3, and 4b are therefore ommited.

```
fullname:  Thomas Cappola
method: Bioconductores 1.2 RMA with VSN instead of quantiles
nickname: RMAVSN
competition: YES
URL: http://www.bioconductor.org

Overall signal to ratio assessment:

           slope        R2   medianSD Null FC IQR Null FC 99.9% Rank
RMAVSN 0.1205668 0.8852778 0.02403352  0.04448677     0.1481711   22

Expected Rank (out of 12626) for genes with fold change of 2
when all other genes are not differentially expressed
stratified by nominal concentration.

           RMAVSN
0.125:0       848
0.25:0.125   2483
0.5:0.25     1688
1:0.5          38
2:1             4
4:2             4
8:4            16
16:8           30
32:16          12
```

```
64:32           8
128:64         10
256:128        79
512:256       465
```

Table 1 for Spike-in Data:

	RMAVSN
Signal detect slope	0.12056681
Signal detect R2	0.88527781
AUC (FP<10)	0.43211399
AUC (FP<15)	0.47608691
AUC (FP<25)	0.53061976
AUC (FP<100)	0.68912502
AFP, call if fc>2	0.05494505
ATP, call if fc>2	7.68498168
IQR	0.11897620
Obs-intended-fc slope	0.12023148
Obs-(low)int-fc slope	0.07888583
FC=2, AUC (FP<10)	0.14852608
FC=2, AUC (FP<15)	0.17578057
FC=2, AUC (FP<25)	0.20668628
FC=2, AUC (FP<100)	0.29481416
FC=2, AFP, call if fc>2	0.00000000
FC=2, ATP, call if fc>2	2.90476190

Table 2 for Spike-in Data:

	RMAVSN
null log-fc IQR	0.04448677
null log-fc 99%	0.11072339
null log-fc 99.9%	0.14817107
low AUC	0.34664466
med AUC	0.40431007
high AUC	0.19207397
weighted avg AUC	0.35893866
25% SD	0.01679203
Median SD	0.02403352
75% SD	0.03249221
99% SD	0.09755627
low.slope	0.08084467
med.slope	0.12423121
high.slope	0.10057811

low.R2	0.22896738
med.R2	0.50578165
high.R2	0.64316311
0.125:0	0.05230988
0.25:0.125	0.02640037
0.5:0.25	0.03615787
1:0.5	0.11241136
2:1	0.15496915
4:2	0.15547705
8:4	0.12665837
16:8	0.11606908
32:16	0.13268689
64:32	0.13970402
128:64	0.13625155
256:128	0.09945088
512:256	0.06640762

Figure 1

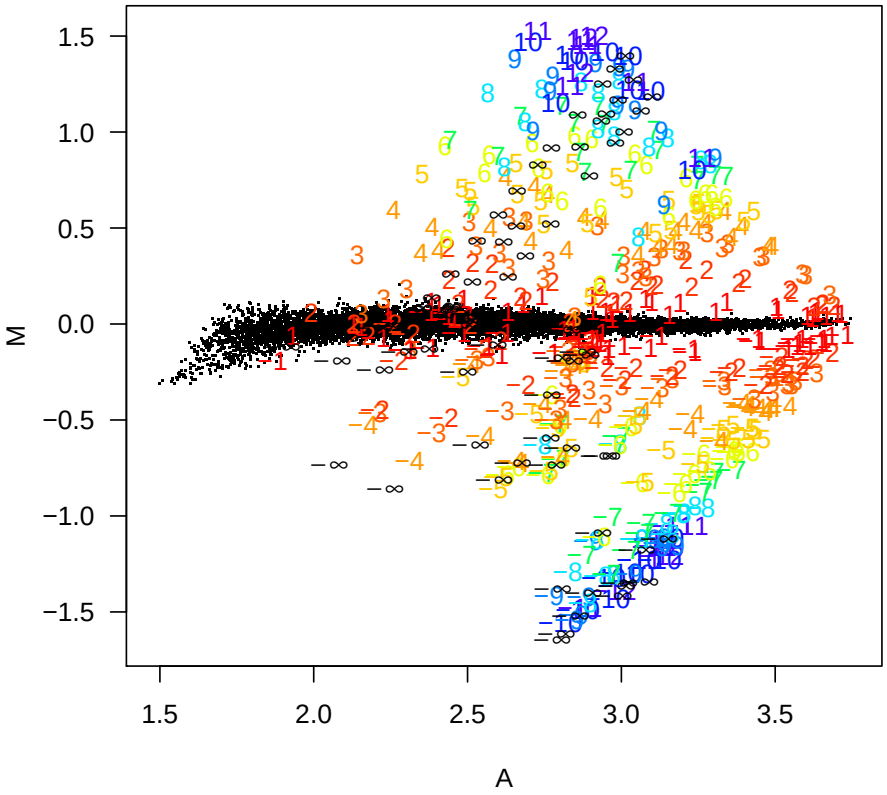

Figure 1: The MA plot shows log fold change as a function of mean log expression level. A set of 14 arrays representing a single experiment from the Affymetrix spike-in data are used for this plot. A total of 13 sets of fold changes are generated by comparing the first array in the set to each of the others. Spiked-in genes are symbolized by numbers representing the nominal $\log_2$ fold change for the gene. Non-differentially expressed genes with observed fold changes larger than 2 are plotted in red. All other probesets are represented with black dots.

Figure 1b

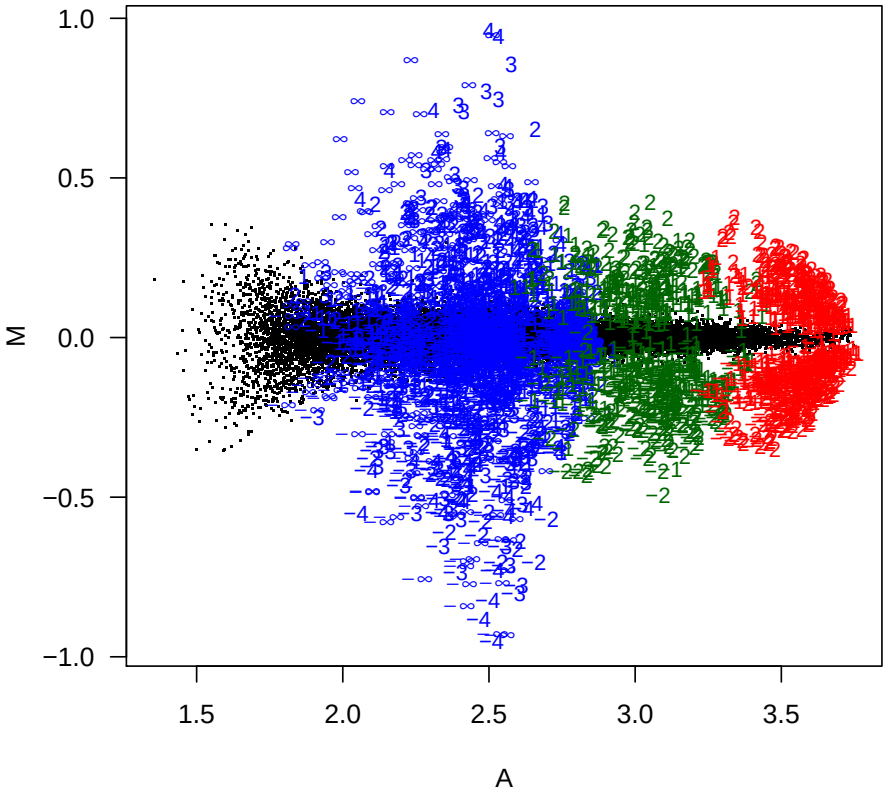

Figure 1b: The MA plot shows log fold change as a function of mean log expression level. A set of 28 arrays representing a single experiment from the Affymetrix spike-in data are used for this plot. Fold changes are generated for all possible comparisons of the the first 14 arrays and the second 14 arrays. Spiked-in genes are symbolized by numbers representing the nominal $\log_2$ fold change for the gene. Of the genes that are spiked to be differentially expressed, only genes with small nominal fold changes are shown. The colors represent four different groups: nominal concentration of genes being compared less than or equal to 2 picoMolar (blue), between 4 and 32 picoMolar (green), greater than or equal to 64 picoMolar (blue). Non-differentially expressed genes with observed fold changes larger than 2 are plotted in red. All other probesets are represented with black dots.

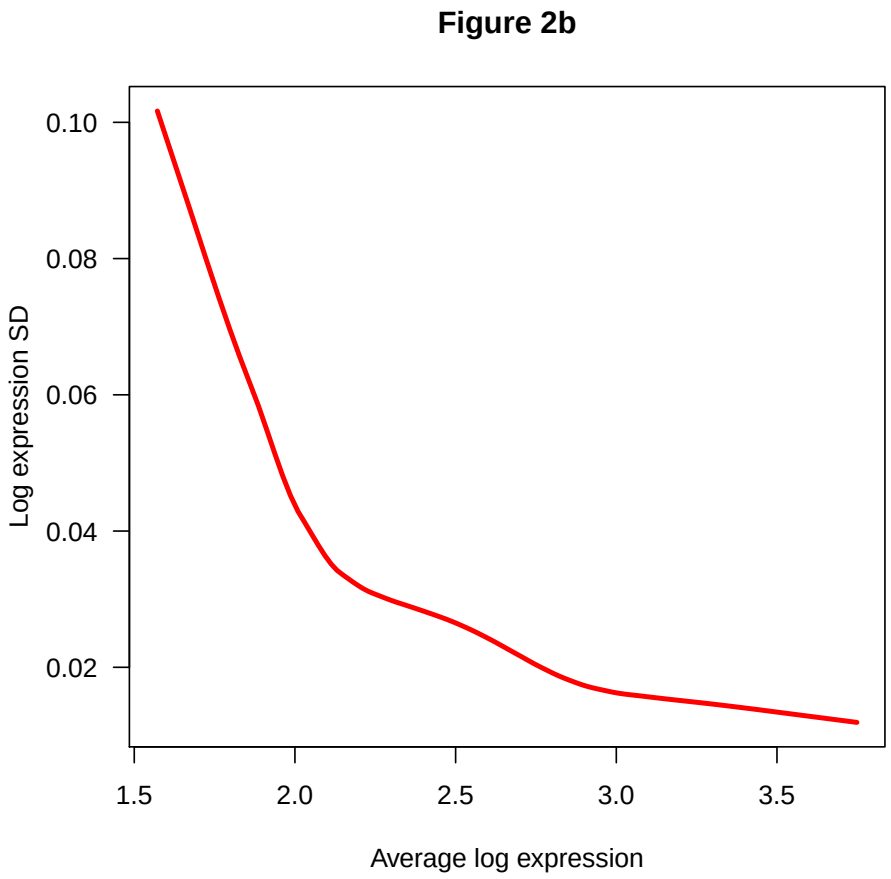

Figure 2b: For each non-spiked-in gene in the 28 arrays used in Figure 1b, we calculate the mean log expression and the observed standard deviation across the 28 replicates. The resulting scatterplot is smoothed to generate a single curve representing mean standard deviation as a function of mean log expression.

Figure 4a

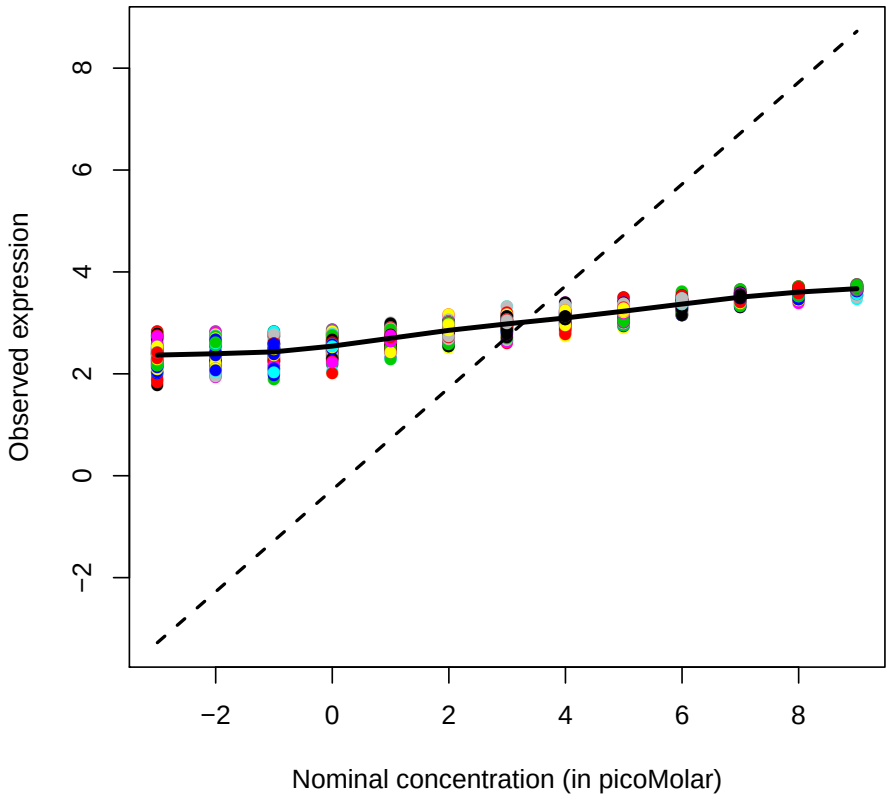

Figure 4a) Average observed log_2 intensity plotted against nominal log_2 concentration for each spiked-in gene for all arrays in Affymetrix spike-In experiment. The dashed line has the ideal slope of 1.

Figure 4c

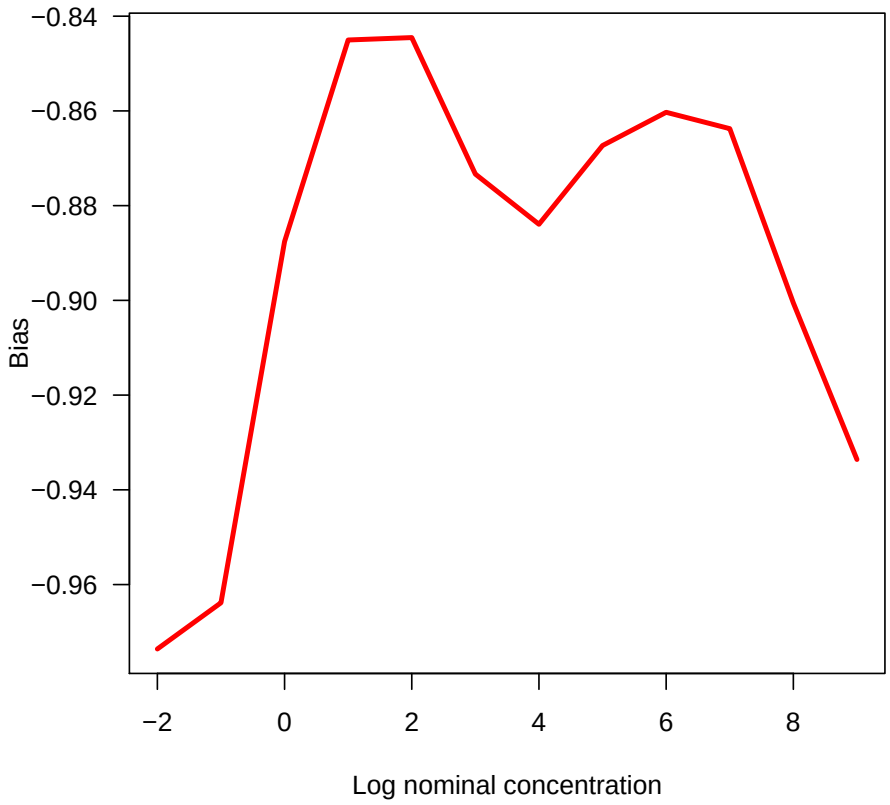

Figure 4c) Using the 28 arrays of Figure 1b, we compute local slopes. As the slopes shown in Figure 4a), the local slopes represent the expected observed log fold-change for probesets with true fold-change of 2 but they are presented as a function of the total nominal probeset concentration in the two samples being compared. In theory the local slopes should be one so we show the bias (difference between the observed local slope and one).

Figure 5a

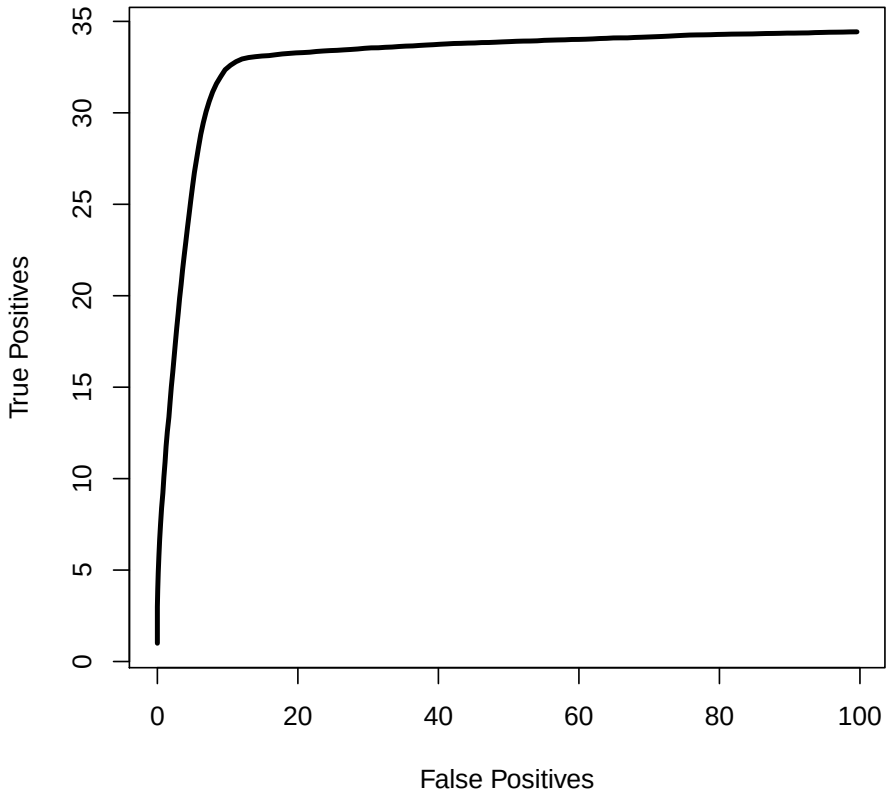

Figure 5a) A typical identification rule for differential expression filters genes with fold change exceeding a given threshold. This figure shows average ROC curves which offer a graphical representation of both specificity and sensitivity for such a detection rule. Average ROC curves based on comparisons with nominal fold changes ranging from 2 to 4096.

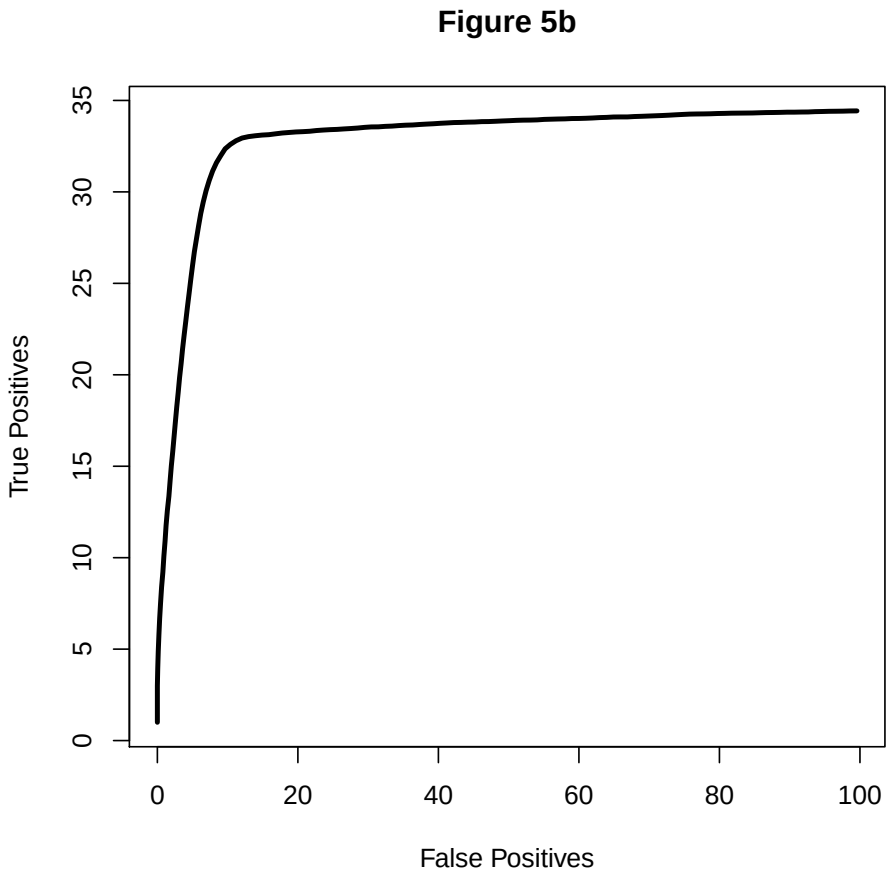

Figure 5b) As 5a) but with nominal fold changes equal to 2.

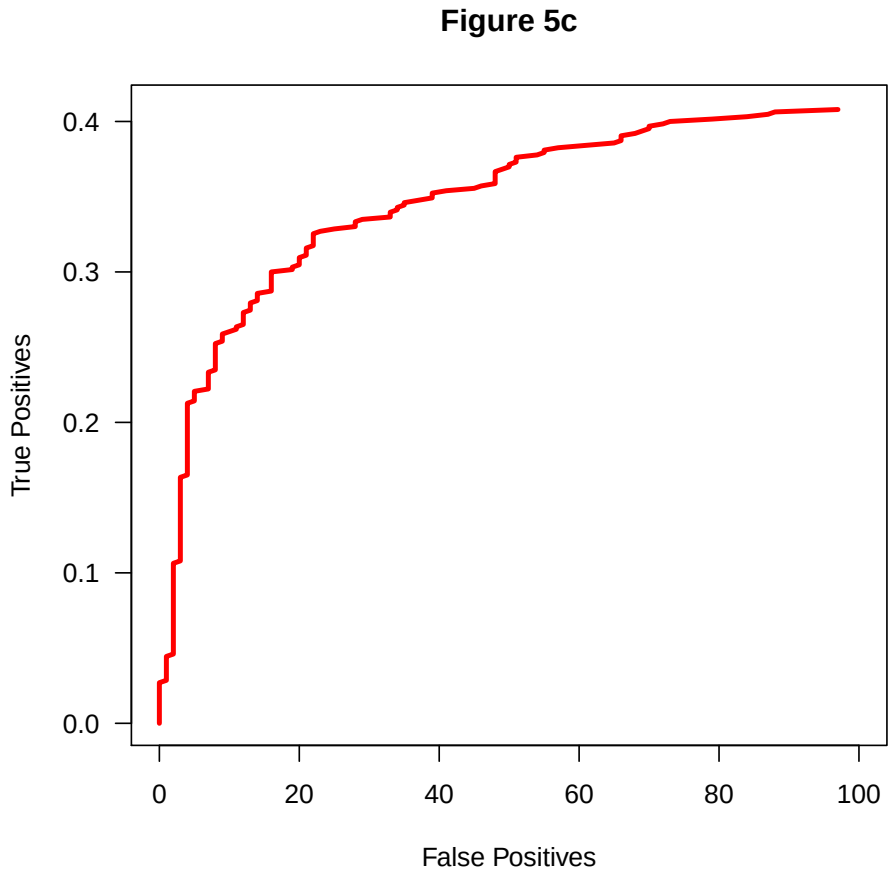

Figure 5c) As 5a) but for comparisons with both nominal concentrations at most 4 picoMolar and nominal fold changes at most 2.

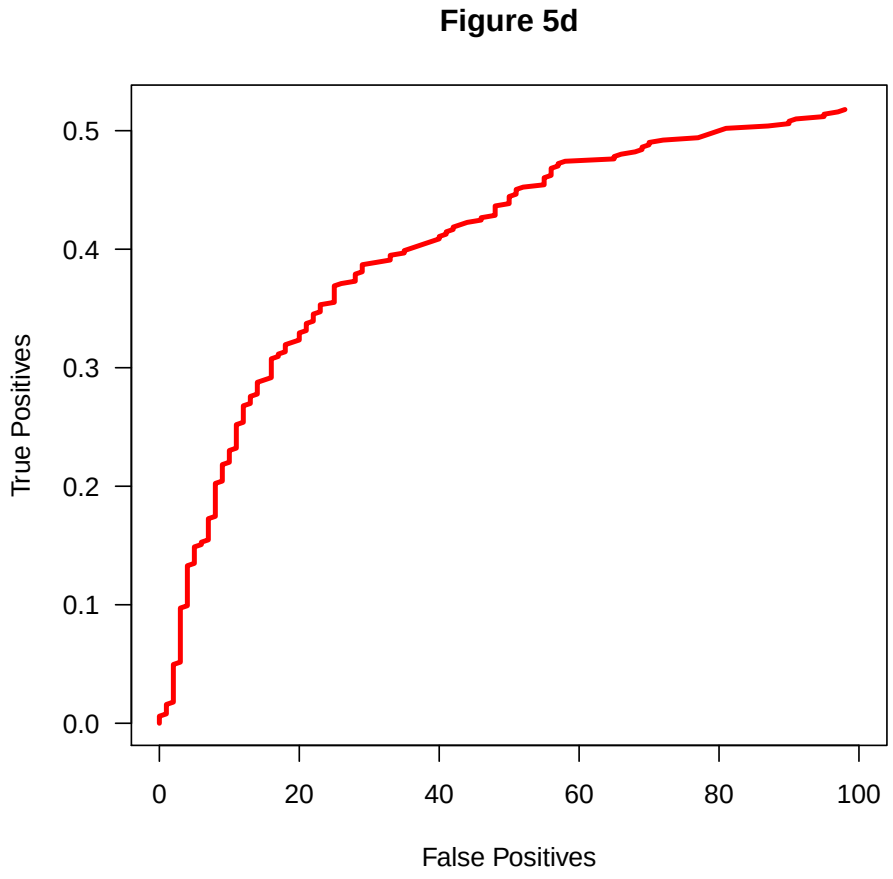

Figure 5d) As 5a) but for comparisons with both nominal concentrations between 4 and 64 picoMolar and nominal fold changes at most 2.

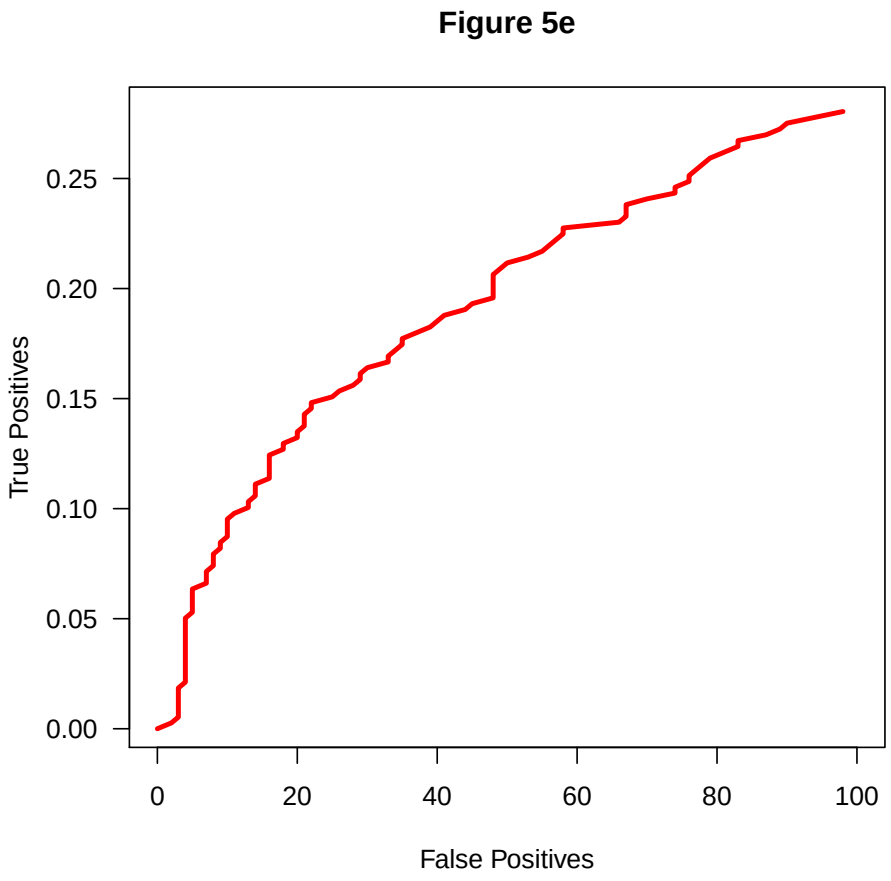

Figure 5e) As 5a) but for comparisons with both nominal concentrations at least 64 and with nominal fold changes at most 2.

Figure 6a

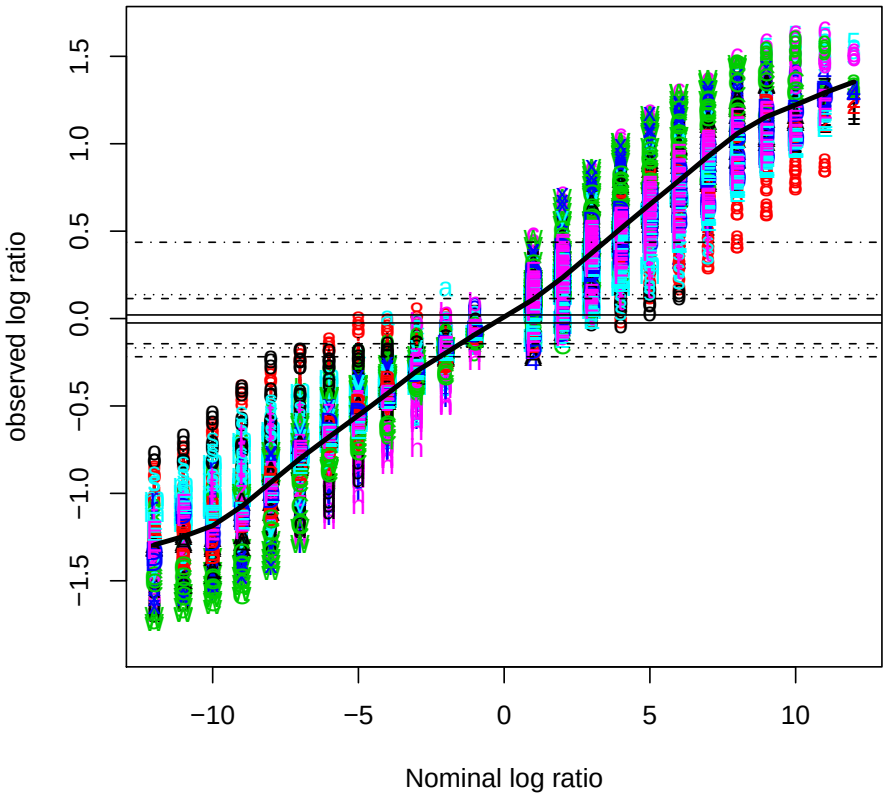

Figure 6a) Observed log fold changes plotted against nominal log fold changes. The dashed lines represent highest, 25th highest, 100th highest, 25th percentile, 75th percentile, smallest 100th, smallest 25th, and smallest log fold change for the genes that were not differentially expressed.

Figure 6b

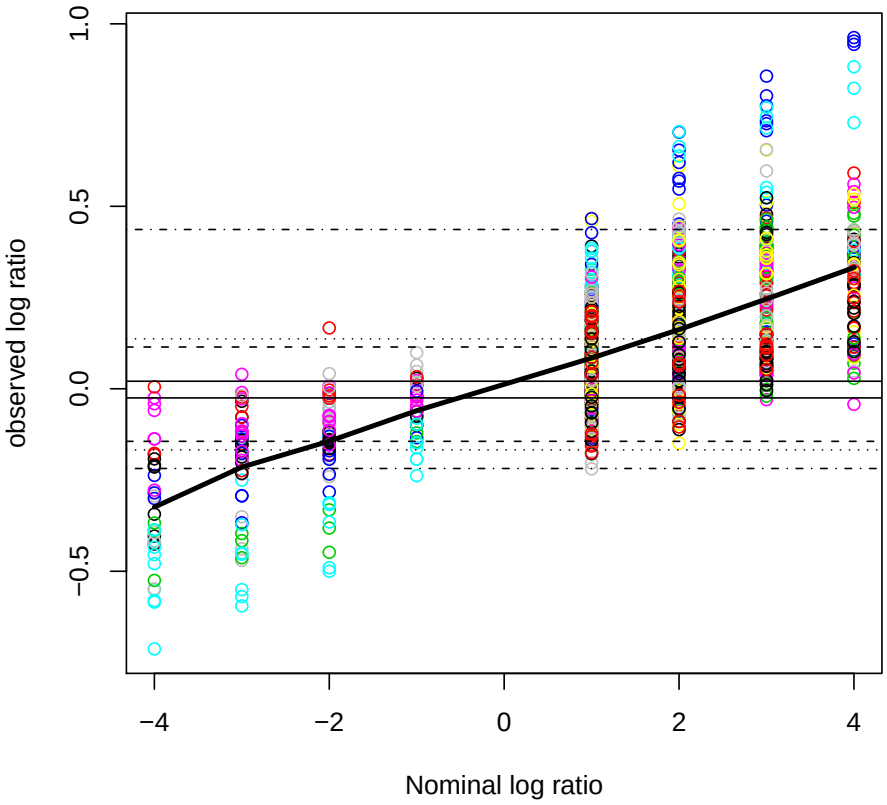

Figure 6b) Like a) but the observed fold changes were calculated for spiked in genes with nominal concentrations no higher than 2pM.