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: Xuemei Feng
method: GCRMA
nickname: GCRMA
competition: YES

Overall signal to ratio assessment:

          slope        R2   medianSD Null FC IQR Null FC 99.9% Rank
GCRMA 0.9929221 0.9148337 0.03552734           0     0.6405807    1

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

           GCRMA
0.125:0     1060
0.25:0.125  1235
0.5:0.25    1294
1:0.5          7
2:1            2
4:2            2
8:4            2
16:8           2
32:16          2
64:32          2
```

```
128:64         2
256:128        2
512:256        5
```

Table 1 for Spike-in Data:

	GCRMA
Signal detect slope	0.9929221
Signal detect R2	0.9148337
AUC (FP<10)	0.5202115
AUC (FP<15)	0.5553375
AUC (FP<25)	0.6091924
AUC (FP<100)	0.7675686
AFP, call if fc>2	3.4432234
ATP, call if fc>2	35.3553114
IQR	0.4103747
Obs-intended-fc slope	0.9915009
Obs-(low)int-fc slope	0.4891927
FC=2, AUC (FP<10)	0.3703272
FC=2, AUC (FP<15)	0.4079420
FC=2, AUC (FP<25)	0.4595979
FC=2, AUC (FP<100)	0.6032705
FC=2, AFP, call if fc>2	1.7619048
FC=2, ATP, call if fc>2	19.1904762

Table 2 for Spike-in Data:

	GCRMA
null log-fc IQR	0.00000000
null log-fc 99%	0.37926588
null log-fc 99.9%	0.64058072
low AUC	0.45307245
med AUC	0.85785007
high AUC	0.85088271
weighted avg AUC	0.55419718
25% SD	0.00000000
Median SD	0.03552734
75% SD	0.08343553
99% SD	0.20000612
low.slope	0.48961001
med.slope	1.06498510
high.slope	0.97995164
low.R2	0.36302407

med.R2	0.45958398
high.R2	0.61634723
0.125:0	0.10620088
0.25:0.125	0.09104557
0.5:0.25	0.08660308
1:0.5	0.64312161
2:1	1.26241743
4:2	1.30425147
8:4	1.16393890
16:8	0.97378915
32:16	1.08762588
64:32	1.13678838
128:64	1.22388646
256:128	1.01218075
512:256	0.69304466

Figure 1

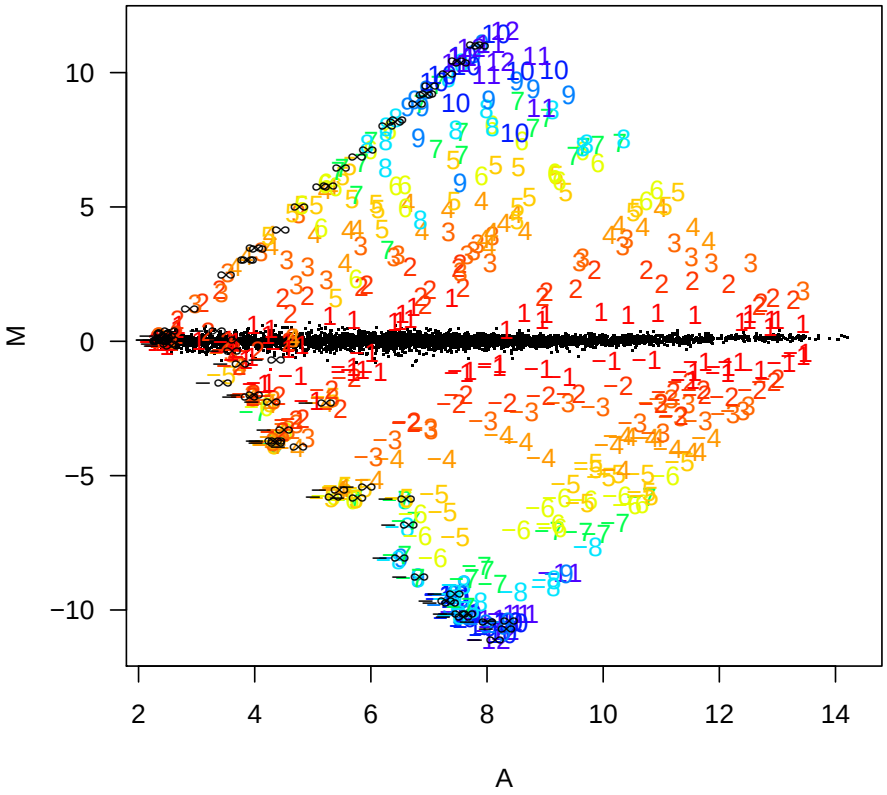

Figure 1: The MA plot shows log fold change as a function of mean log expression level. A set of 14 arrays representing a single experiment from the Affymetrix spike-in data are used for this plot. A total of 13 sets of fold changes are generated by comparing the first array in the set to each of the 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

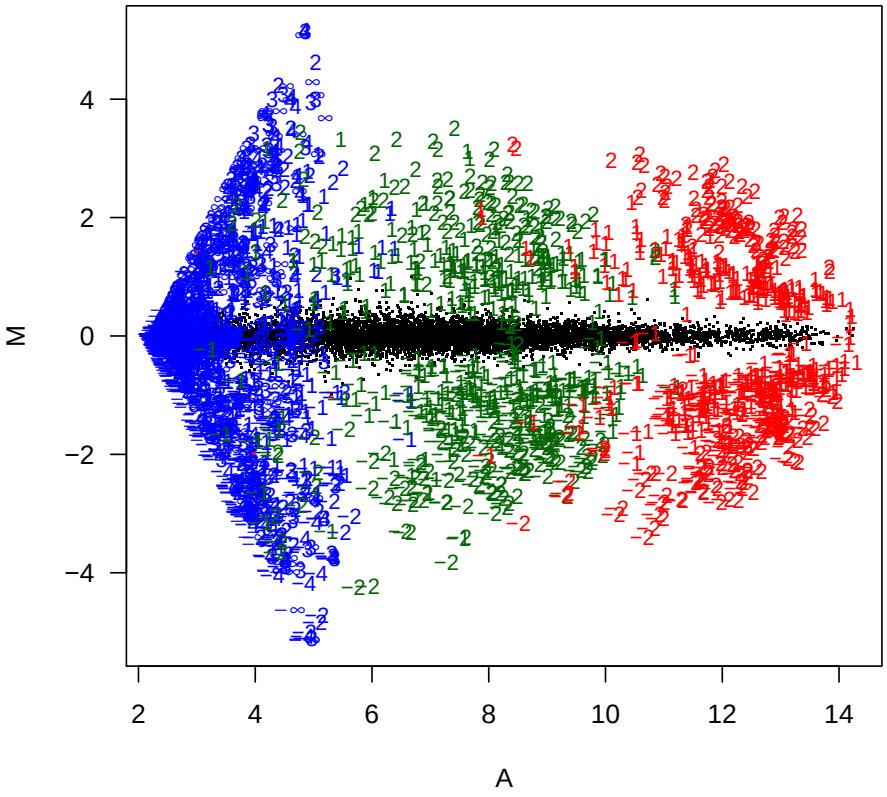

Figure 1b: The MA plot shows log fold change as a function of mean log expression level. A set of 28 arrays representing a single experiment from the Affymetrix spike-in data are used for this plot. Fold changes are generated for all possible comparisons of the the first 14 arrays and the second 14 arrays. 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.

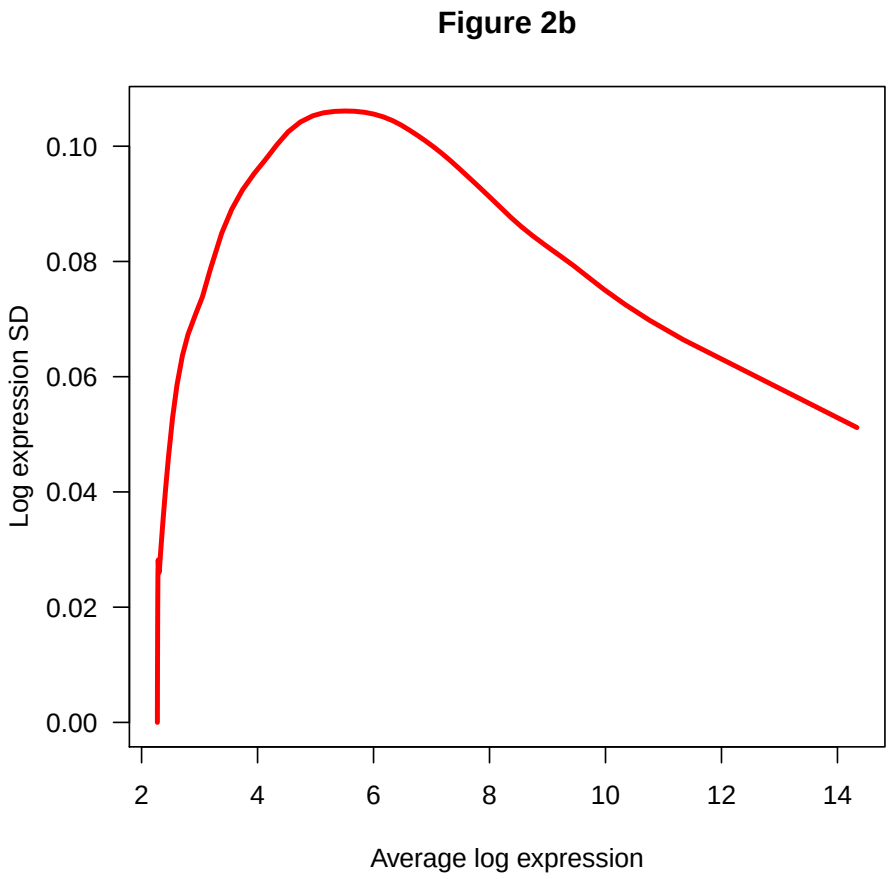

Figure 2b: For each non-spiked-in gene in the 28 arrays used in Figure 1b, we calculate the mean log expression and the observed standard deviation across the 28 replicates. The resulting scatterplot is smoothed to generate a single curve representing mean standard deviation as a function of mean log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

Figure 4a

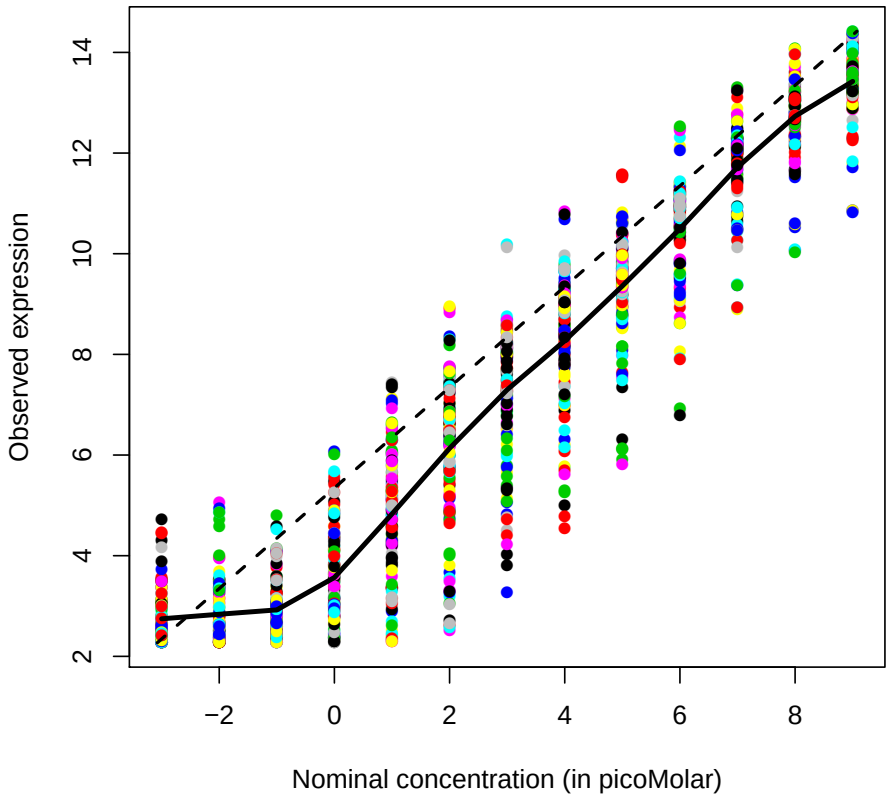

Figure 4a) Average observed log_2 intensity plotted against nominal log_2 concentration for each spiked-in gene for all arrays in Affymetrix spike-In experiment. The dashed line has the ideal slope of 1.

Figure 4c

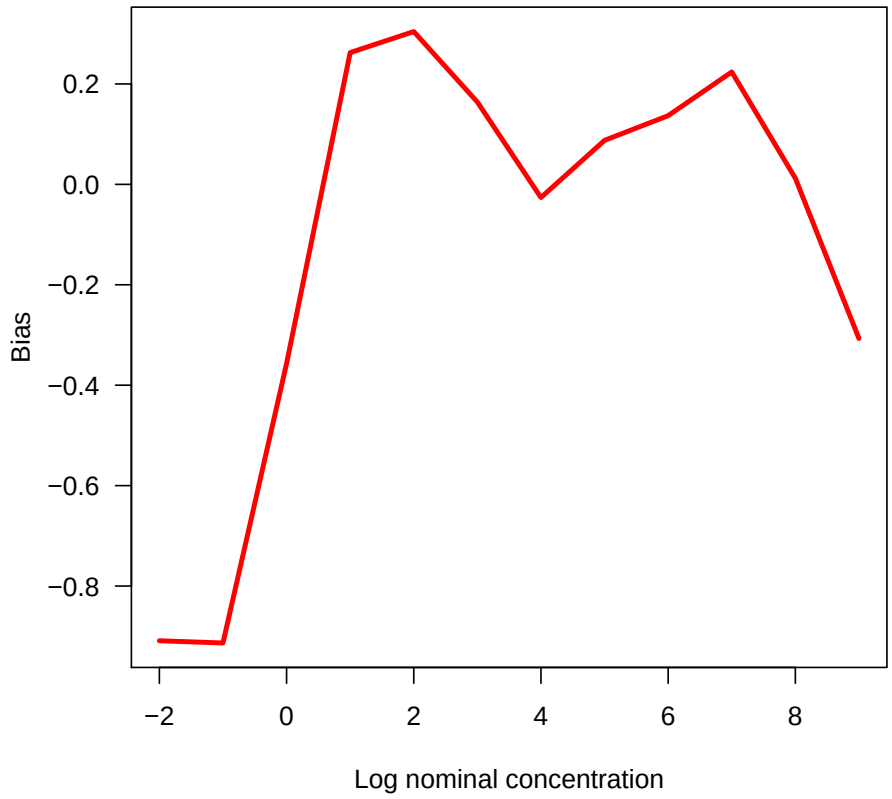

Figure 4c) Using the 28 arrays of Figure 1b, we compute local slopes. As the slopes shown in Figure 4a), the local slopes represent the expected observed log fold-change for probesets with true fold-change of 2 but they are presented as a function of the total nominal probeset concentration in the two 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

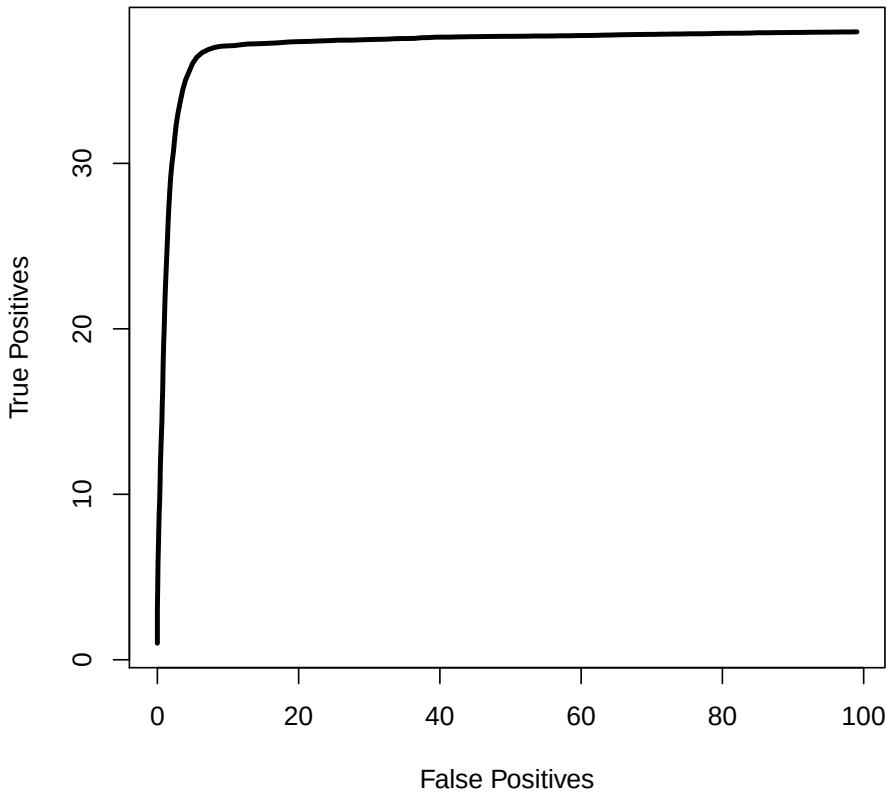

Figure 5a) A typical identification rule for differential expression filters genes with fold change exceeding a given threshold. This figure shows average ROC curves which offer a graphical representation of both specificity and sensitivity for such a detection rule. Average ROC curves based on comparisons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

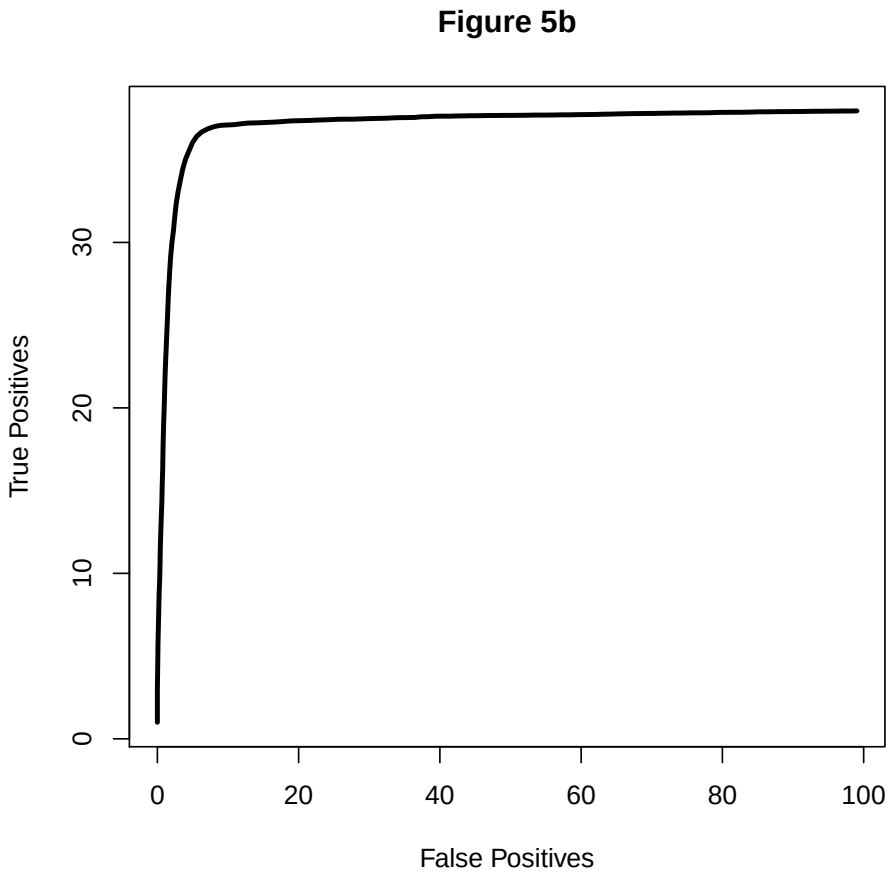

Figure 5b) As 5a) but with nominal fold changes equal to 2.

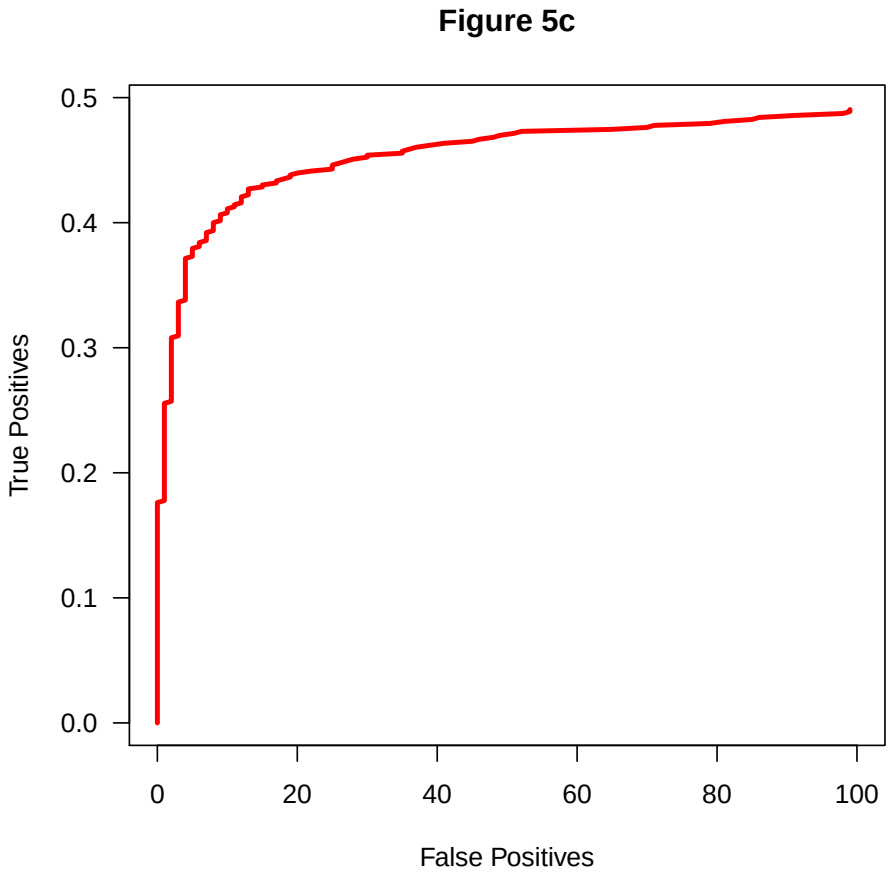

Figure 5c) As 5a) but for comparisons with both nominal concentrations at most 4 picoMolar and nominal fold changes at most 2.

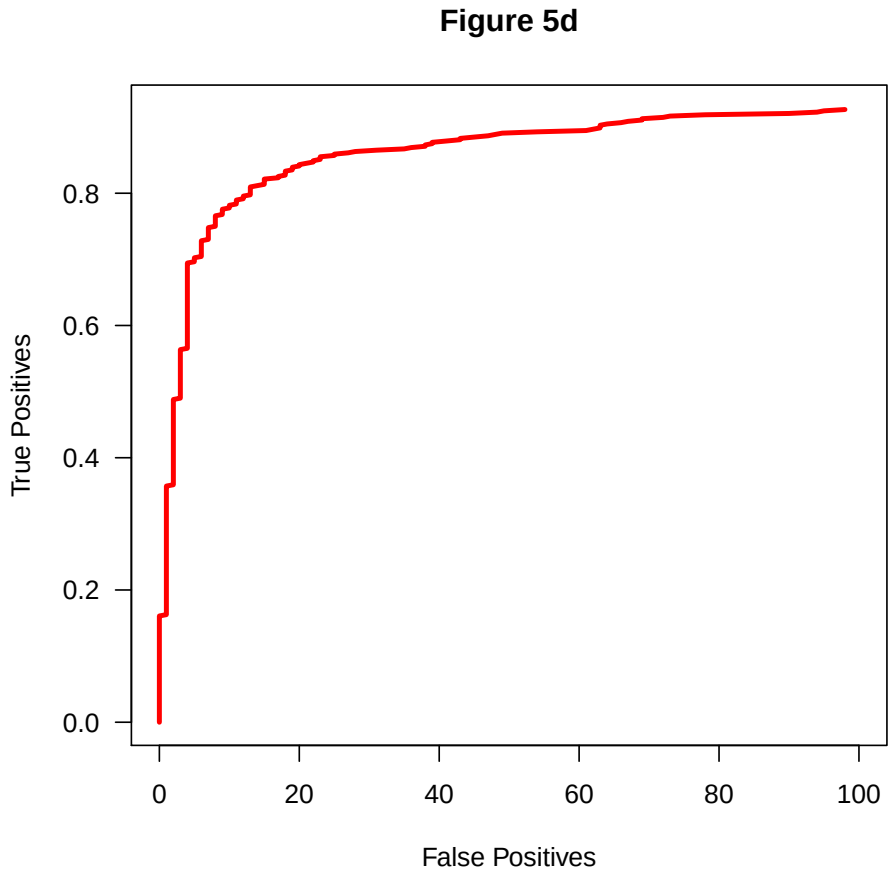

Figure 5d) As 5a) but for comparisons with both nominal concentrations between 4 and 64 picoMolar and nominal fold changes at most 2.

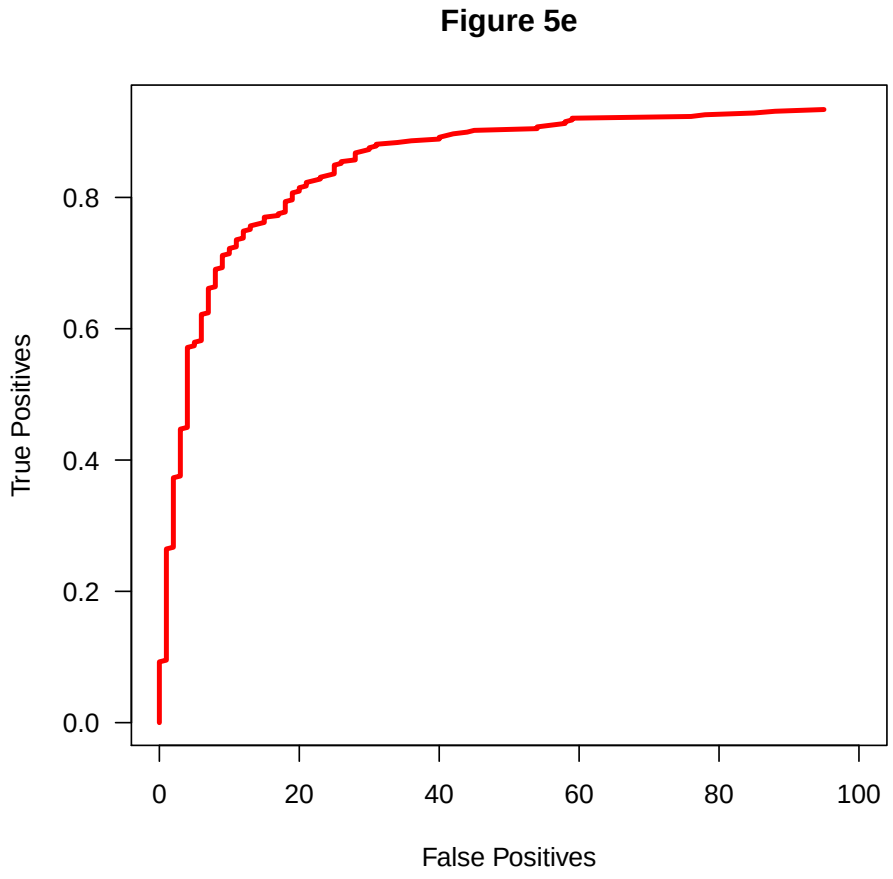

Figure 5e) As 5a) but for comparisons with both nominal concentrations at least 64 and with nominal fold changes at most 2.

Figure 6a

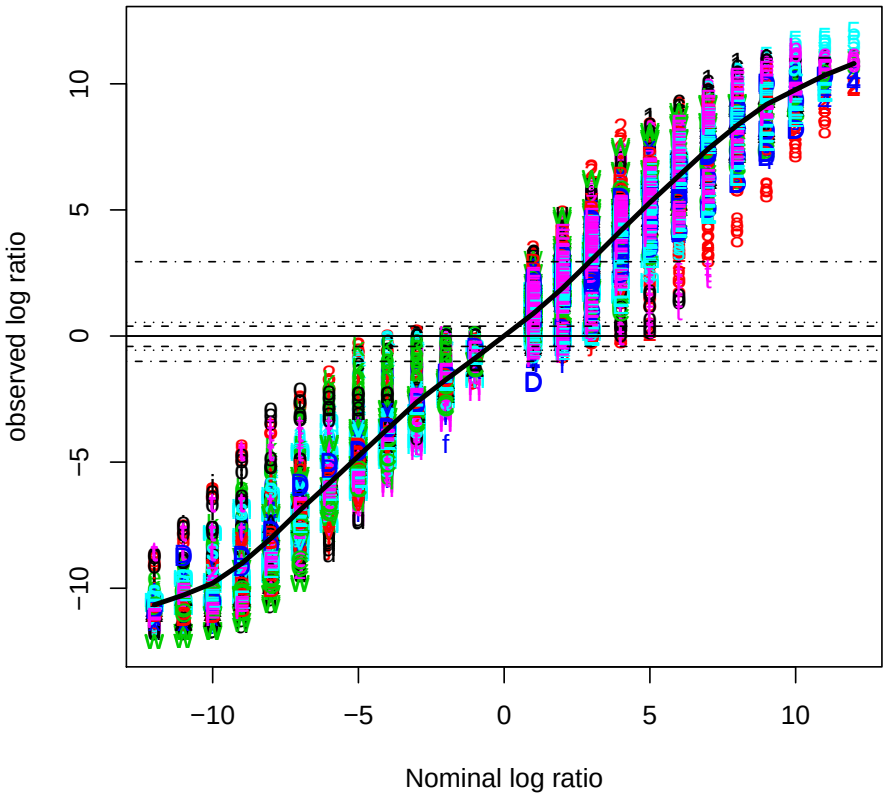

Figure 6a) Observed log fold changes plotted against nominal log fold changes. The dashed lines represent highest, 25th highest, 100th highest, 25th percentile, 75th percentile, smallest 100th, smallest 25th, and smallest log fold change for the genes that were not differentially expressed.

Figure 6b

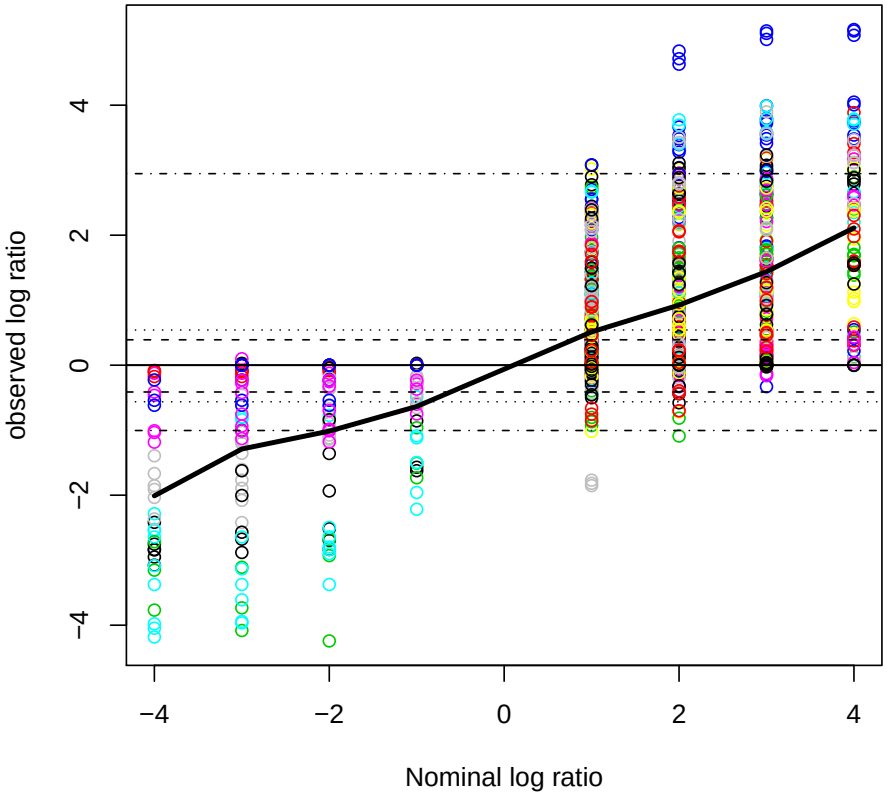

Figure 6b) Like a) but the observed fold changes were calculated for spiked in genes with nominal concentrations no higher than 2pM.