Bioconductor Expression Assessment Tool for Affymetrix Oligonucleotide Arrays (affycomp)

June 8, 2011

Contents

```
fullname: Rafael A. Irizarry
method: MAS 5.0
nickname: MAS 5.0
competition: YES
URL: http://www.affymetrix.com/support/technical/technotes/statistical_algorithms_techn
```

Table:

```
> tableAll(tmp.assessment)
```

	MAS 5.0
Median SD	2.920239e-01
R2	8.890008e-01
1.25v20 corr	7.297434e-01
2-fold discrepancy	1.226000e+03
3-fold discrepancy	3.320000e+02
Median slope	8.474941e-01
Signal detect slope	7.058227e-01
Signal detect R2	8.565416e-01
AUC (FP<10)	2.171241e-01
AUC (FP<15)	2.379589e-01
AUC (FP<25)	2.696288e-01
AUC (FP<100)	3.557402e-01
AFP, call if fc>2	3.107387e+03
ATP, call if fc>2	1.281893e+01
IQR	2.655135e+00
Obs-intended-fc slope	6.932507e-01
Obs-(low)int-fc slope	6.471881e-01

```
FC=2, AUC (FP<10)       6.160714e-02
FC=2, AUC (FP<15)       6.190476e-02
FC=2, AUC (FP<25)       6.232830e-02
FC=2, AUC (FP<100)      6.508575e-02
FC=2, AFP, call if fc>2 3.069786e+03
FC=2, ATP, call if fc>2 3.714286e+00
```

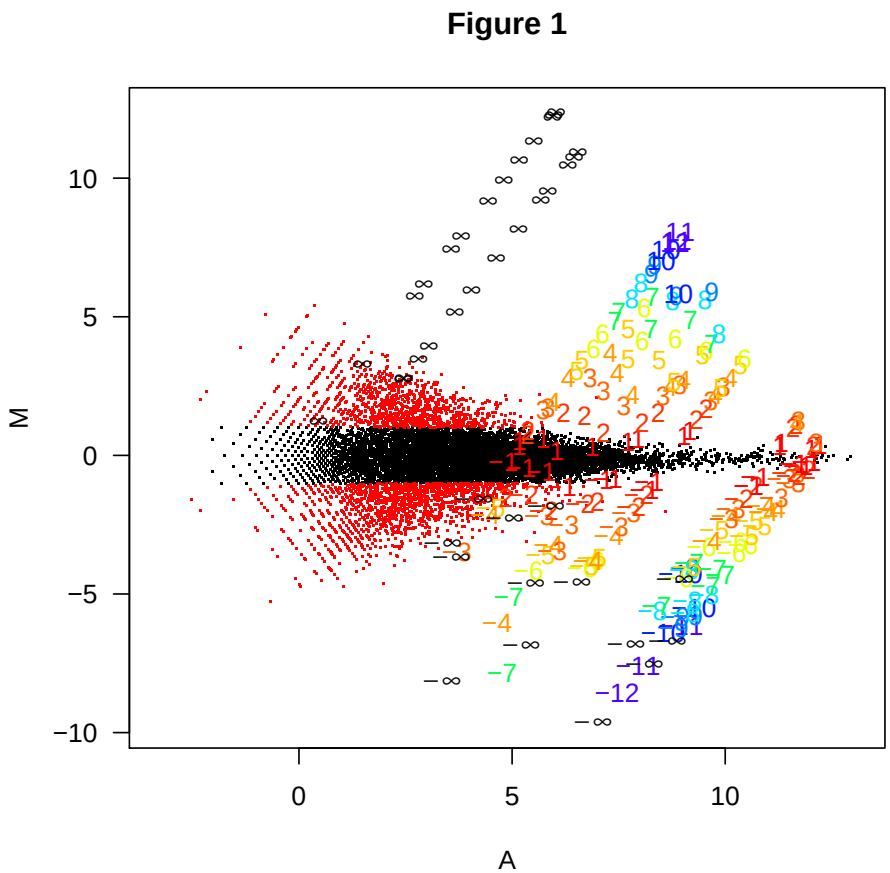

Figure 1: The MA plot shows log fold change as a function of mean log expression level. A set of 14 arrays representing a single experiment from the Affymetrix spike-in data are used for this plot. A total of 13 sets of fold changes are generated by comparing the first array in the set to each of the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

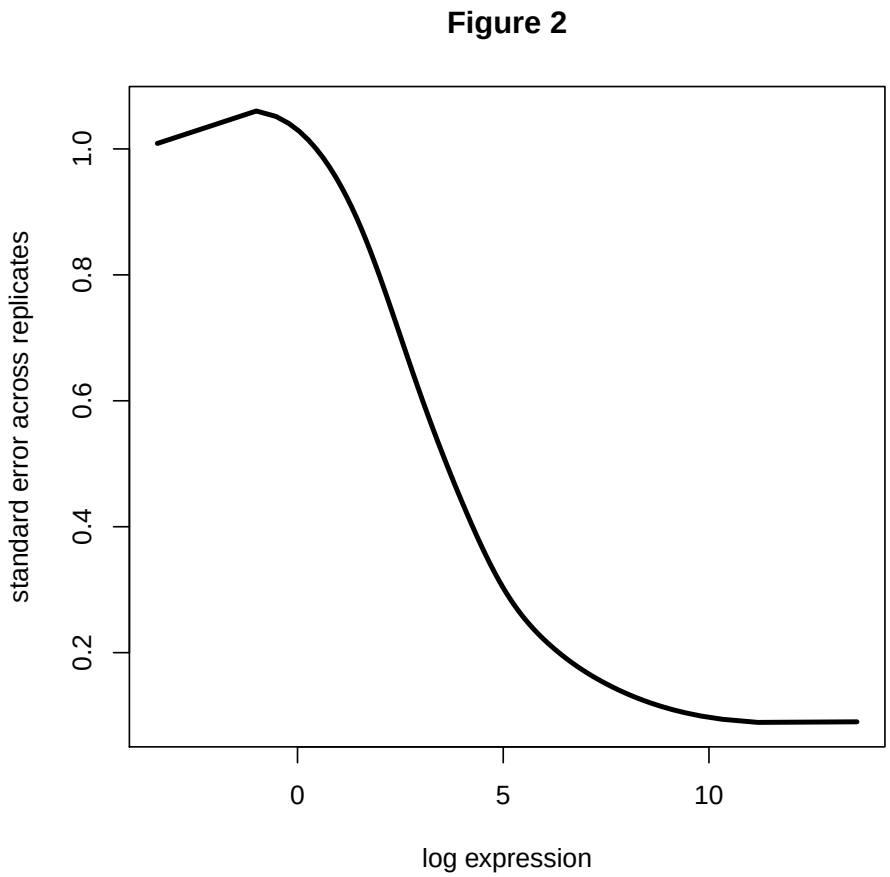

Figure 2: For each gene, and each experimental condition, we calculate the mean log expression and the observed standard deviation across 5 replicates. The resulting scatterplot is smoothed to generate a single curve representing mean standard deviation as a function of mean log expression.

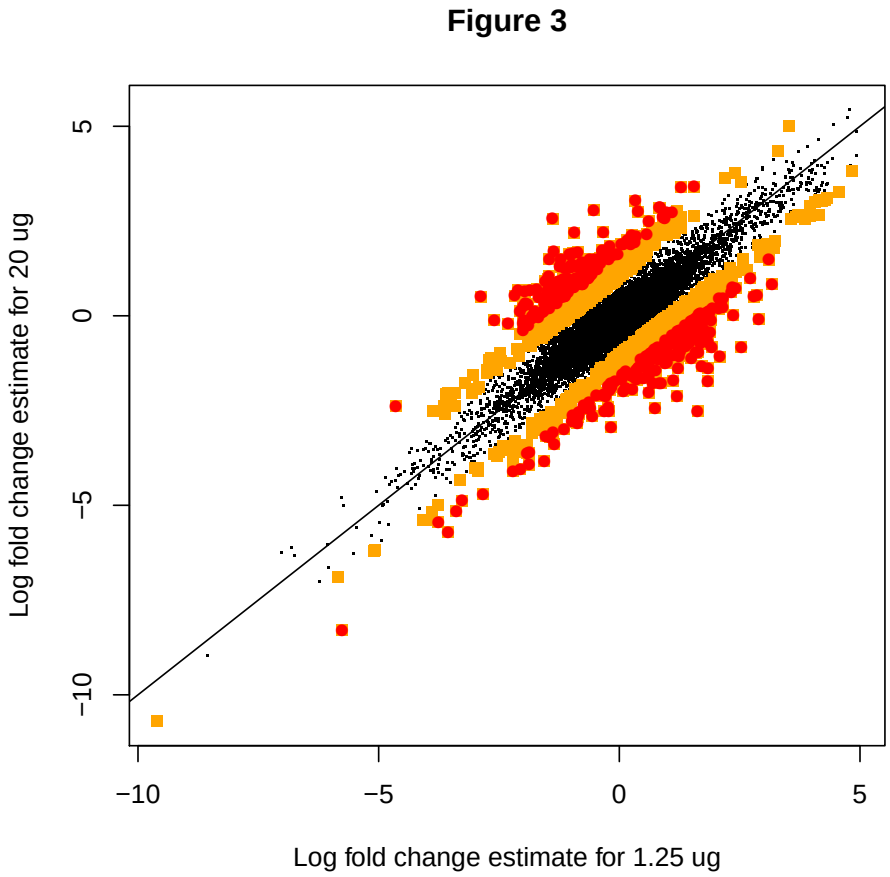

Figure 3: This plot, using the GeneLogic dilution data, shows the sensitivity of fold change calculations to total RNA abundance. Average log fold-changes between liver and CNS for the lowest concentration and the highest in the dilution data set are computed. Orange and red color is used to denote genes with $M_{6g} - M_{1g}$ bigger than $\log_2(2)$ and $\log_2(3)$ respectively. The rest are denoted with black.

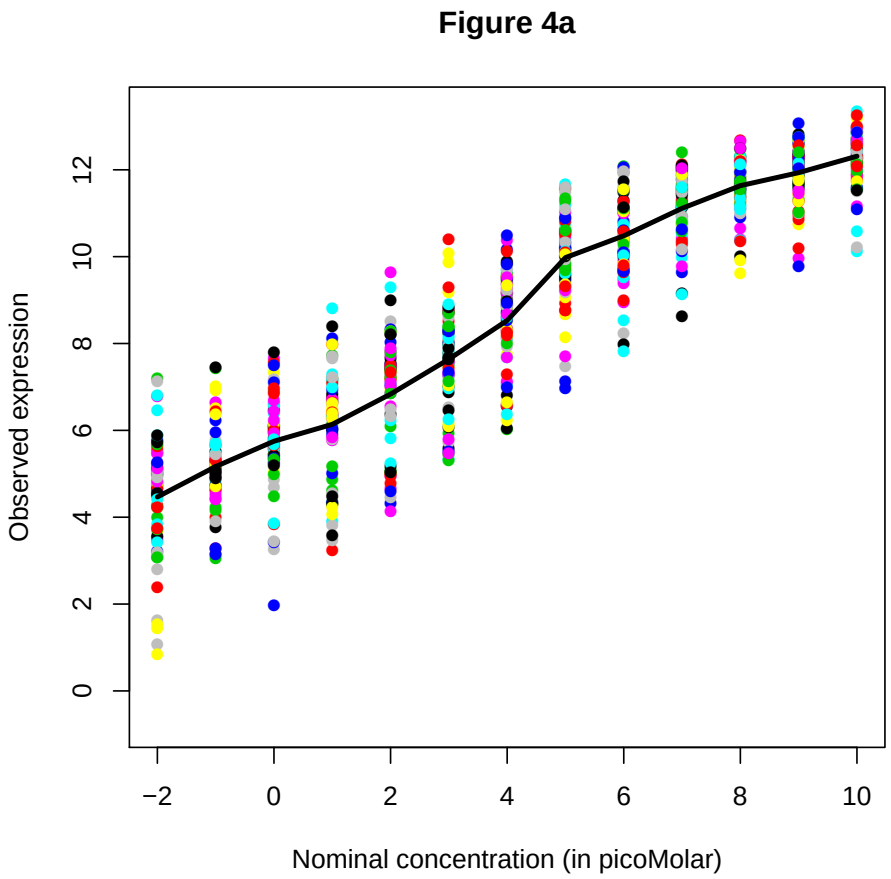

Figure 4a) Average observed log_2 intensity plotted against nominal log_2 concentration for each spiked-in gene for all arrays in Affymetrix spike-In experiment.

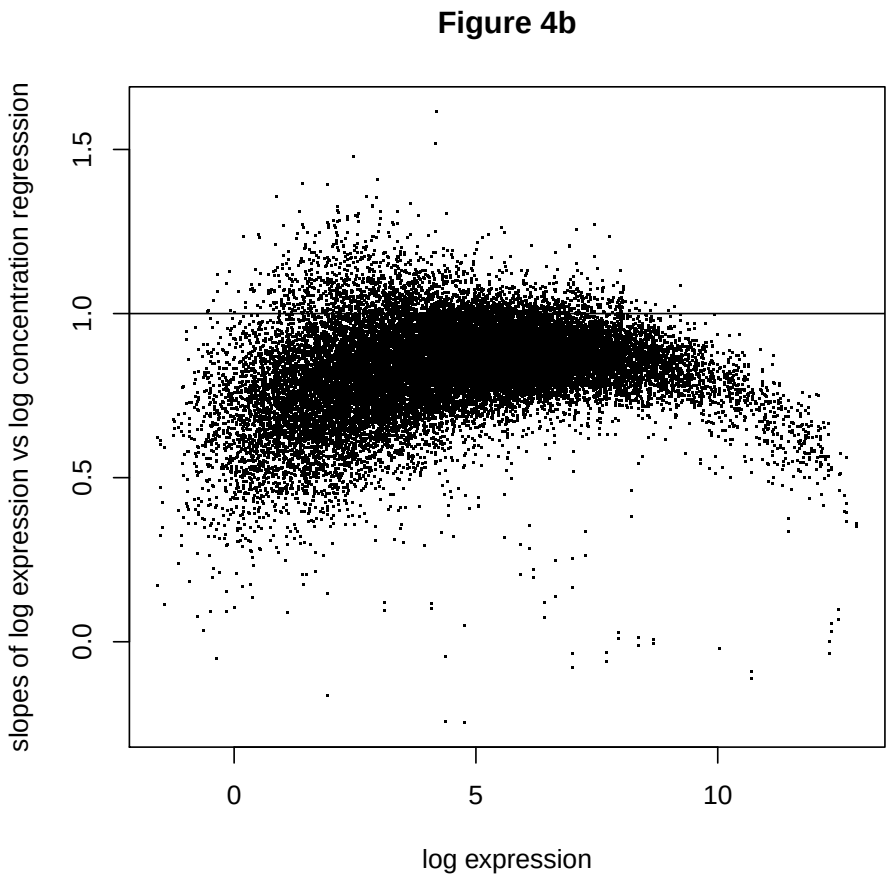

Figure 4b) For the GeneLogic dilution data, log expression values are regressed against their log nominal concentration. The slope estimates are plotted against average log intensity across all concentrations.

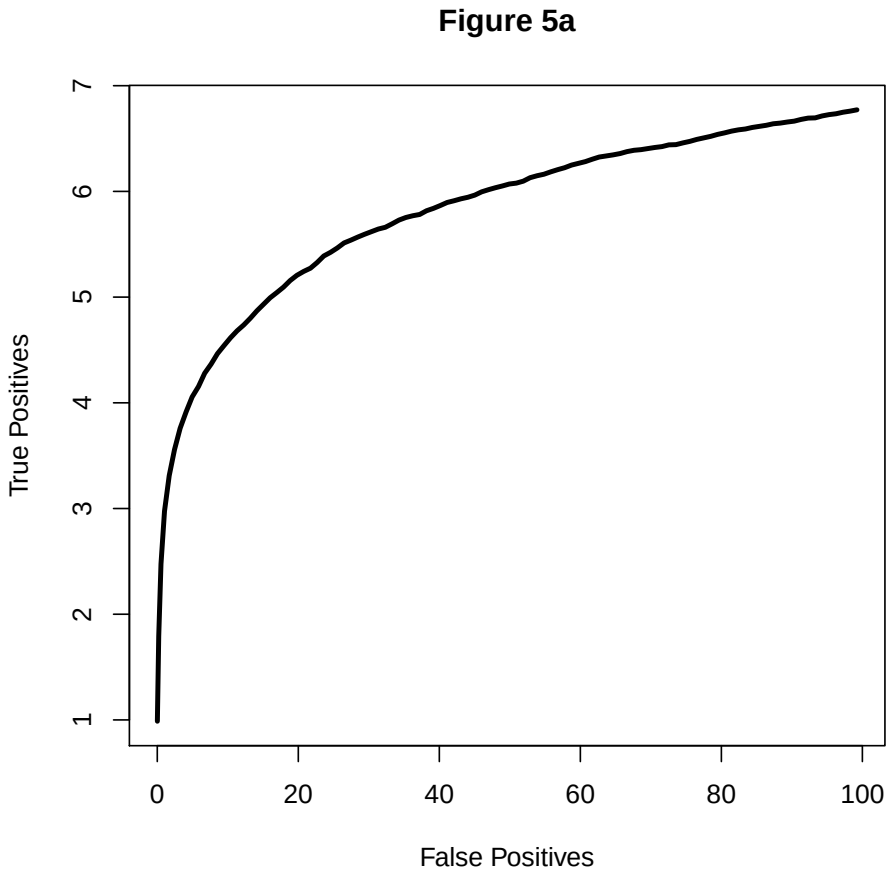

Figure 5a) A typical identification rule for differential expression filters genes with fold change exceeding a given threshold. This figure shows average ROC curves which offer a graphical representation of both specificity and sensitivity for such a detection rule. Average ROC curves based on comparisons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

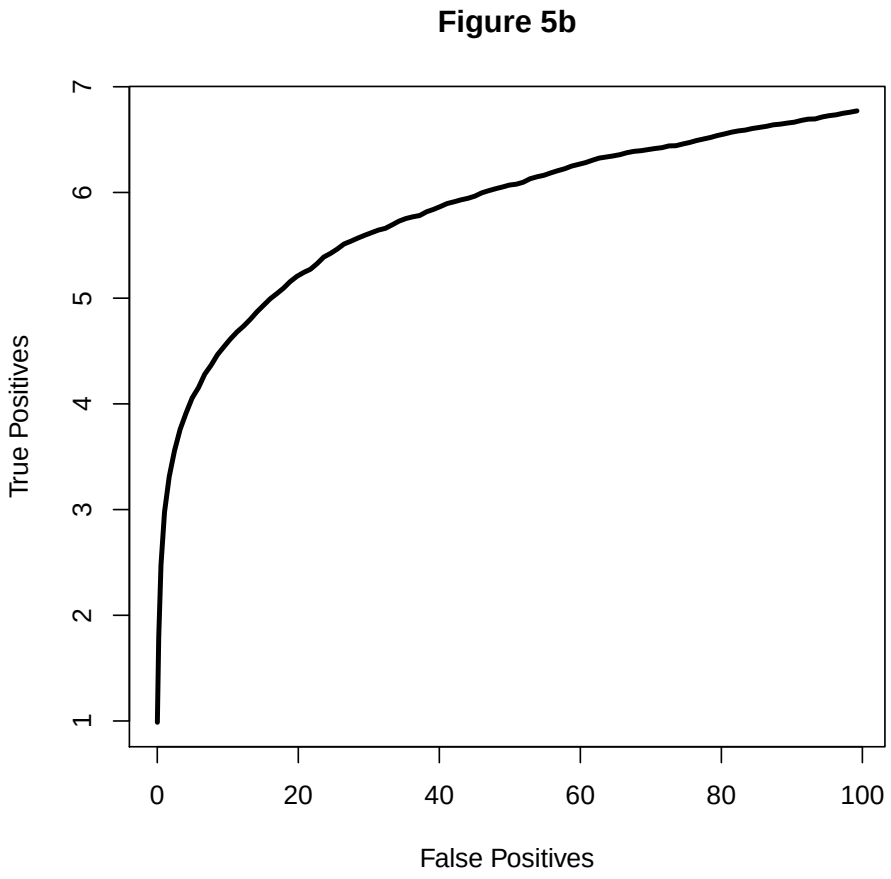

Figure 5b) As 5a) but with nominal fold changes equal to 2.

Figure 6a

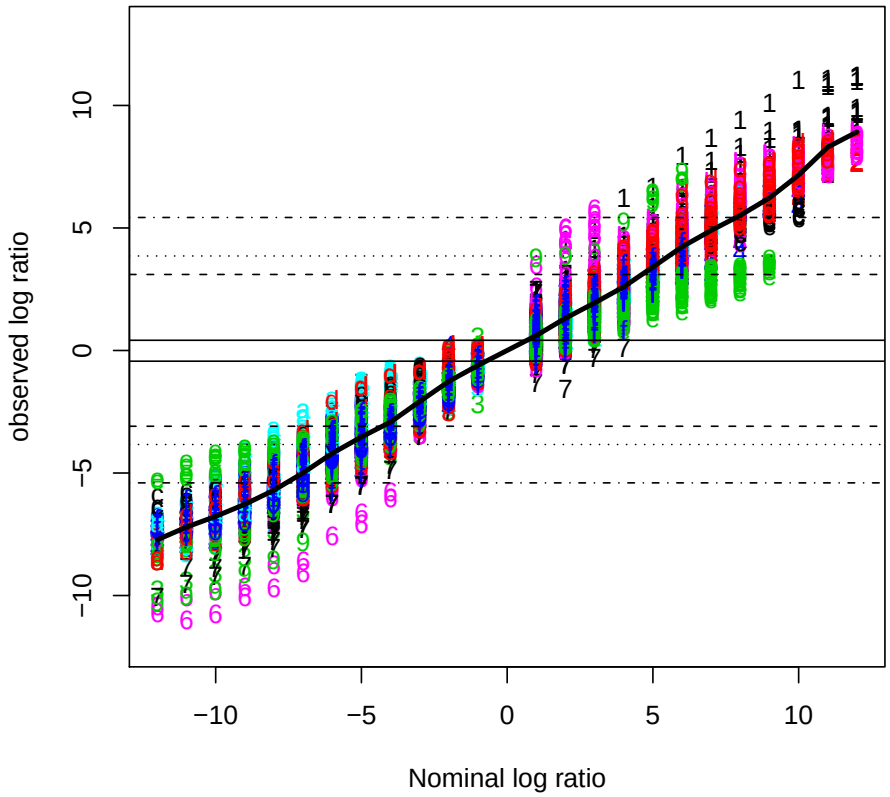

Figure 6a) Observed log fold changes plotted against nominal log fold changes. The dashed lines represent highest, 25th highest, 100th highest, 25th percentile, 75th percentile, smallest 100th, smallest 25th, and smallest log fold change for the genes that were not differentially expressed.

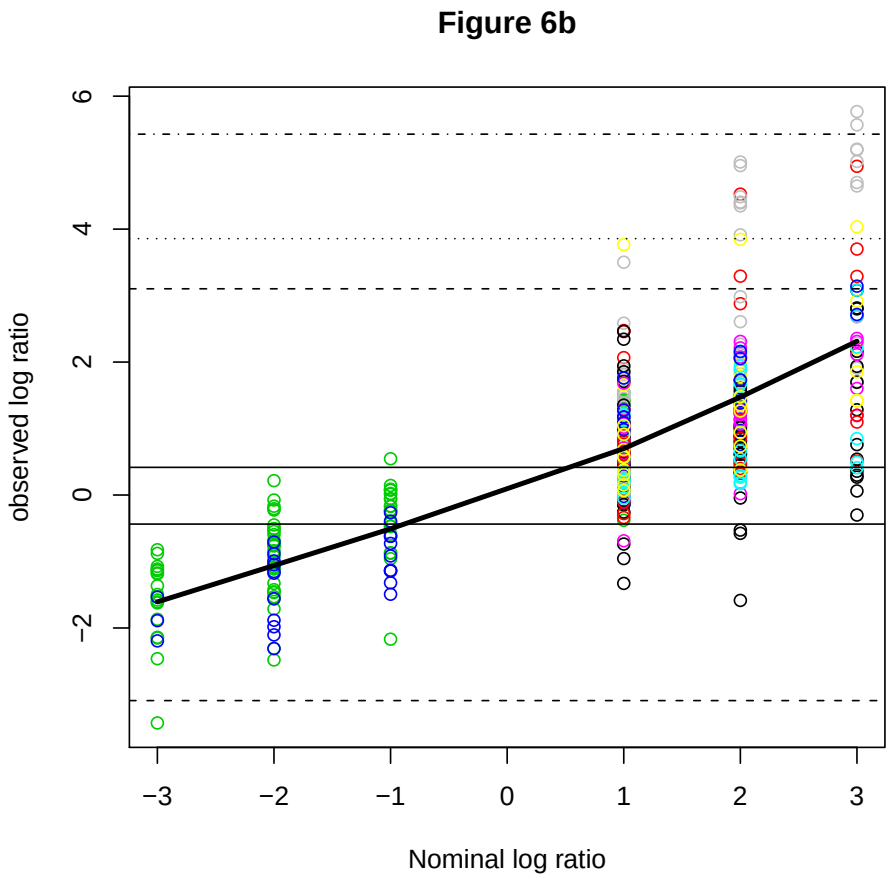

Figure 6b) Like a) but the observed fold changes were calculated for spiked in genes with nominal concentrations no higher than 2pM.